STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 126106

TO: Ginny Portner **Location: REM/3B02/3C18** **Art Unit: 1645** **Tuesday, July 06, 2004** **Case Serial Number: 09/732091**	**From: Toby Port** **Location: Biotech-Chem Library** **Remsen 1A59** **Phone: 571-272-2523** **toby.port@uspto.gov**

Search Notes

Dear Examiner Portner,

Here are the results of your search.
Please feel free to contact me if you have any questions.

Toby Port

STIC-Biotech/ChemLib 10/6/06

From: Portner, Ginny
Sent: Tuesday, June 29, 2004 6:07 PM
To: STIC-Biotech/ChemLib
Subject: RE: 09/732,091

try 09/732,091

Ginny Portner
Remsen Building
Art Unit 1645
Room E03, B02
(571) 272-0862

-----Original Message-----
From: STIC-Biotech/ChemLib
Sent: Tuesday, June 29, 2004 4:23 PM
To: Portner, Ginny
Subject: RE: 09/732,081

There is no valid CRF for this Serial Number. Please provide us with another Serial Number from the Parent Application.
LEONARD 22520

-----Original Message-----
From: Portner, Ginny
Sent: Tuesday, June 29, 2004 4:21 PM
To: STIC-Biotech/ChemLib
Subject: 09/732,081
Importance: High

Please search SEQ ID NO 3 and 4 with respect to polypeptides/proteins. The nucleic acid should be back translated to the polypeptide. Thanks.

Ginny Portner
Remsen Building
Art Unit 1645
Room E03, B02
(571) 272-0862

Searcher:_____
Phone:_____
Location:_____
Date Picked Up:_____
Date Completed:_____
Searcher Prep/Review:_____
Clerical:_____
Online time:_____

TYPE OF SEARCH:
NA Sequences:_____
AA Sequences:_____
Structures:_____
Bibliographic:_____
Litigation:_____
Full text:_____
Patent Family:_____
Other:_____

VENDOR/COST (where applic.)
STN:_____
DIALOG:_____
Questel/Orbit:_____
DRLink:_____
Lexis/Nexis:_____
Sequence Sys.:_____
WWW/Internet:_____
Other (specify):_____

```
CC   Helicobacter sp. HP30 protein
XX
SQ   Sequence 253 AA;

  Query Match             100.0%;  Score 1279;  DB 5;  Length 253;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-118;
  Matches  253;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              |||||||||||||
Db        241 ANGDKKSLQIESI 253


RESULT 2
AAE26878
ID   AAE26878 standard; protein; 265 AA.
XX
AC   AAE26878;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Helicobacter sp. PQE/HP30 protein.
XX
KW   HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW   type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW   low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW   immunostimulant; cytostatic; peptic ulcer.
XX
OS   Helicobacter sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 12
FT                   /note= "Encoded by TTC"
XX
PN   WO200251237-A2.
XX
PD   04-JUL-2002.
XX
PF   07-DEC-2001; 2001WO-US048392.
XX
PR   07-DEC-2000; 2000US-00732091.
XX
PA   (ANTE-) ANTEX BIOLOGICS INC.
XX
PI   Tian J,  Walker R,  Jackson WJ;
XX
DR   WPI; 2002-666854/71.
DR   N-PSDB; AAD44535.
XX
PT   Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT   the proteins, useful as vaccines for raising immune response in animals.
XX
PS   Disclosure; Page 122-123; 127pp; English.
XX
CC   The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC   peptides derived from them. The invention is useful for producing an
CC   immune response. It is useful for preventing, treating or ameliorating a
CC   disorder or disease associated with infection of an animal with
CC   Helicobacter. Pharmaceutical composition and vaccines comprising the
CC   sequences of the invention is useful for treating type B gastritis,
CC   peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC   cell lymphoma. The invention is useful as reagents for clinical or
CC   medical diagnosis of Helicobacter infections and for scientific research
CC   on the properties of pathogenicity, virulence and infectivity of
CC   Helicobacter, as well as host defense mechanisms. The present sequence is
CC   Helicobacter sp. PQE/HP30 protein
XX
SQ   Sequence 265 AA;

  Query Match             100.0%;  Score 1279;  DB 5;  Length 265;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-118;
  Matches  253;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 72

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 132

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 192

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 252

Qy        241 ANGDKKSLQIESI 253
              |||||||||||||
Db        253 ANGDKKSLQIESI 265


RESULT 3
AAW20486
ID   AAW20486 standard; protein; 253 AA.
XX
AC   AAW20486;
XX
DT   29-JUL-1997  (first entry)
XX
DE   H. pylori cytoplasmic protein, 4095342.aa.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS   Helicobacter pylori.
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US009122.
XX
PR   07-JUN-1995;   95US-00487032.
PR   01-APR-1996;   96US-00630405.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Smith D,  Berglindh OT,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   N-PSDB; AAT67811.
XX
PT   Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT   useful for vaccines to treat or prevent H. pylori infection, and to
PT   detect Helicobacter.
XX
```

PS Claim 61; Page 651; 1481pp; English.
XX
CC The present sequence is a H. pylori cytoplasmic protein. The protein may
CC be used in a vaccine to prevent or treat H. pylori infection or to
CC identify H. pylori polypeptide binding compounds, useful as potential H.
CC pylori life cycle activators or inhibitors. The genomic sequence of H.
CC pylori (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The sequences were analysed for
CC ORF of at least 180 nucleotides, and the predicted coding regions defined
CC by computer evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF were
CC analysed for significant homology to other known or exported membrane
CC proteins. Having identified and determined the sequences of interest,
CC particular regions can be isolated from H. pylori by PCR amplification
CC for recombinant polypeptide production, e.g. in E. coli hosts
XX
SQ Sequence 253 AA;

Query Match 99.3%; Score 1270; DB 2; Length 253;
Best Local Similarity 99.2%; Pred. No. 1.4e-117;
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              || |||||||||:
Db        241 ANEDKKSLQIESV 253
```

RESULT 4
AAW24673
ID AAW24673 standard; protein; 253 AA.
XX
AC AAW24673;
XX
DT 12-AUG-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 4095342.aa.
XX
KW Transmembrane; cytoplasmic; cell envelope; flagella; transport; secreted;
KW periplasmic; chronic gastritis; duodenal ulcer disease; activator;
KW inhibitor; bacterial life cycle; vaccine; immunise; detection; antisense;
KW inhibition.
XX
OS Helicobacter pylori.
XX
PN WO9719098-A1.
XX
PD 29-MAY-1997.
XX
PF 15-NOV-1996; 96WO-US018542.
XX
PR 17-NOV-1995; 95US-00561469.
XX
PA (ASTR) ASTRA AB.
XX
PI Smith DH;
XX
DR WPI; 1997-298052/27.
DR N-PSDB; AAT77491.
XX
PT Helicobacter pylori nucleic acid sequences and related proteins - used
PT for diagnostics and therapeutics.
XX
PS Claim 18; Page 184; 235pp; English.
XX
CC This sequence represents an H. pylori cytoplasmic protein. Helicobacter
CC pylori has been strongly linked to chronic gastritis and duodenal ulcer
CC disease. The nucleic acid sequences of the invention are used to evaluate
CC compounds, especially activators or inhibitors of bacterial life cycle,
CC for the ability to bind an H. pylori nucleic acid sequence. The nucleic
CC acid sequences, and corresponding proteins, are also useful for
CC generating vaccines for immunising subjects against H. pylori or for use
CC in detecting the presence of Helicobacter species in a sample. Antisense
CC nucleic acid sequences of these sequences are used to inhibit expression
CC of a gene from Helicobacter species. H. pylori whole genomic DNA was
CC isolated and nebulised to a median size of 2000 bp. Purified DNA
CC fragments were blunt-ended and ligated to unique BstXI-linker adapters in
CC 100-1000 fold molar excess. These linkers are complementary to the BstXI-
CC cut pMPX vectors, while the overhang is not self-complementary. Therefore
CC the linkers will not concatemerise nor will the cut vector re-ligate
CC itself easily. The linker-adapter inserts were ligated to each of the 20
CC pMPX vectors to construct a series of shotgun subclone libraries. The
CC purified DNA samples were then sequenced. Note: The ORF/protein reference
CC number for this sequence was obtained from the related specification,
CC WO9640893
XX
SQ Sequence 253 AA;

Query Match 99.3%; Score 1270; DB 2; Length 253;
Best Local Similarity 99.2%; Pred. No. 1.4e-117;
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              || |||||||||:
Db        241 ANEDKKSLQIESV 253
```

RESULT 5
AAW20866
ID AAW20866 standard; protein; 256 AA.
XX
AC AAW20866;
XX
DT 18-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 12ge20305orf30.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis.
XX
OS Helicobacter pylori.
XX
PN WO9640893-A1.

```
Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
Db         13 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 32

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
Db         33 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 52

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
Db         53 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 72

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
Db         73 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 92

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
Db         93 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 112

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
Db        113 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 132

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
Db        133 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 152

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
Db        153 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 172

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
Db        173 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 192

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
Db        193 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 212

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
Db        213 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 232

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
Db        233 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 252

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
Db        253 AlaAsnGlyAspLysLysSerLeuGlnIleGluSerIle 265
```

RESULT 3
AAW20486
ID AAW20486 standard; protein; 253 AA.
XX
AC AAW20486;
XX
DT 29-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 4095342.aa.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US009122.
XX
PR 07-JUN-1995; 95US-00487032.
PR 01-APR-1996; 96US-00630405.
XX
PA (ASTR) ASTRA AB.
XX
PI Smith D, Berglindh OT, Mellgaerd BL;
XX
DR WPI; 1997-052306/05.
DR N-PSDB; AAT67811.
XX
PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT useful for vaccines to treat or prevent H. pylori infection, and to
PT detect Helicobacter.
XX
PS Claim 61; Page 651; 1481pp; English.
XX
CC The present sequence is a H. pylori cytoplasmic protein. The protein may
CC be used in a vaccine to prevent or treat H. pylori infection or to
CC identify H. pylori polypeptide binding compounds, useful as potential H.
CC pylori life cycle activators or inhibitors. The genomic sequence of H.
CC pylori (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The sequences were analysed for
CC ORF of at least 180 nucleotides, and the predicted coding regions defined
CC by computer evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF were
CC analysed for significant homology to other known or exported membrane
CC proteins. Having identified and determined the sequences of interest,
CC particular regions can be isolated from H. pylori by PCR amplification
CC for recombinant polypeptide production, e.g. in E. coli hosts
XX
SQ Sequence 253 AA;

Alignment Scores:

Pred. No.:	6.94e-134	Length:	253
Score:	1270.00	Matches:	251
Percent Similarity:	99.60%	Conservative:	1
Best Local Similarity:	99.21%	Mismatches:	1
Query Match:	94.14%	Indels:	0
DB:	2	Gaps:	0

US-09-732-091-3 (1-759) x AAW20486 (1-253)

```
Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
Db         41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
Db         61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
Db         81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
Db        101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
Db        121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
```

GenCore version 5.1.6

CM protein - protein search, using sw model

Run on: July 5, 2004, 03:11:28 ; Search time 34 Seconds
(without alignments)
2316.318 Million cell updates/sec

Title: US-09-732-091-4
Perfect score: 1279
Sequence: 1 MAYKYDRDLEFLKQLESSDL.....LRLKTQQANGDKKSLQIESI 253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1279	100.0	253	9	US-09-732-091-4	Sequence 4, Appli
2	1279	100.0	265	9	US-09-732-091-44	Sequence 44, Appl
3	1270	99.3	253	12	US-10-335-977-9162	Sequence 9162, Ap
4	1270	99.3	253	12	US-10-335-977-9163	Sequence 9163, Ap
5	1270	99.3	256	12	US-10-335-977-9164	Sequence 9164, Ap
6	722	56.5	248	12	US-10-335-977-7699	Sequence 7699, Ap
7	716	56.0	237	12	US-10-335-977-7698	Sequence 7698, Ap
8	472.5	36.9	155	10	US-09-882-227-414	Sequence 414, App
9	252	19.7	49	9	US-09-732-091-20	Sequence 20, Appl
10	198	15.5	38	9	US-09-732-091-17	Sequence 17, Appl
11	194	15.2	41	9	US-09-732-091-19	Sequence 19, Appl
12	152	11.9	30	9	US-09-732-091-18	Sequence 18, Appl
13	150	11.7	30	9	US-09-732-091-16	Sequence 16, Appl
14	98.5	7.7	1009	12	US-10-282-122A-43832	Sequence 43832, A
15	97.5	7.6	815	16	US-10-437-963-191043	Sequence 191043,
16	97	7.6	916	12	US-10-282-122A-76490	Sequence 76490, A
17	96.5	7.5	373	15	US-10-369-493-42	Sequence 42, Appl
18	96.5	7.5	701	14	US-10-032-585-7400	Sequence 7400, Ap
19	96.5	7.5	1102	14	US-10-156-761-14395	Sequence 14395, A
20	95	7.4	517	12	US-10-282-122A-54543	Sequence 54543, A
21	95	7.4	2125	9	US-09-919-172-29	Sequence 29, Appl
22	95	7.4	2649	14	US-10-205-219-169	Sequence 169, App
23	95	7.4	2649	15	US-10-341-434-220	Sequence 220, App
24	95	7.4	2649	15	US-10-341-434-230	Sequence 230, App
25	93.5	7.3	1196	12	US-10-282-122A-52737	Sequence 52737, A
26	93	7.3	610	12	US-10-282-122A-47217	Sequence 47217, A
27	92.5	7.2	430	15	US-10-369-493-5229	Sequence 5229, Ap
28	92.5	7.2	996	9	US-09-815-242-5251	Sequence 5251, Ap
29	92.5	7.2	1009	9	US-09-815-242-12141	Sequence 12141, A
30	92	7.2	889	16	US-10-437-963-197045	Sequence 197045,
31	91.5	7.2	241	12	US-10-412-699B-682	Sequence 682, App
32	91.5	7.2	1847	15	US-10-369-493-1075	Sequence 1075, Ap
33	90.5	7.1	659	12	US-10-424-599-222059	Sequence 222059,
34	90.5	7.1	1163	15	US-10-452-024-107	Sequence 107, App
35	90	7.0	573	12	US-10-425-114-43207	Sequence 43207, A
36	90	7.0	744	16	US-10-437-963-122954	Sequence 122954,
37	90	7.0	820	12	US-10-221-278-590	Sequence 590, App
38	90	7.0	820	15	US-10-291-172-590	Sequence 590, App
39	90	7.0	1089	14	US-10-032-585-7664	Sequence 7664, Ap
40	89.5	7.0	281	15	US-10-289-762-749	Sequence 749, App
41	89.5	7.0	560	14	US-10-253-904-50	Sequence 50, Appl
42	89.5	7.0	869	12	US-10-282-122A-58561	Sequence 58561, A
43	89	7.0	384	12	US-10-260-708-60	Sequence 60, Appl
44	88.5	6.9	482	12	US-10-424-599-199420	Sequence 199420,
45	88.5	6.9	552	9	US-09-817-764-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-732-091-4
; Sequence 4, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; TITLE OF INVENTION: thereof
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-4

Query Match 100.0%; Score 1279; DB 9; Length 253;
Best Local Similarity 100.0%; Pred. No. 4.8e-115;
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSPASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSPASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              |||||||||||||
Db        241 ANGDKKSLQIESI 253


RESULT 2
US-09-732-091-44
; Sequence 44, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
;  APPLICANT:  Tian, Jing-Hui
;  APPLICANT:  Walker, Richard I.
;  APPLICANT:  Jackson, W. James
;  TITLE OF INVENTION:  Helicobacter proteins, gene sequences and uses
;  TITLE OF INVENTION:  thereof
;  FILE REFERENCE:  7969-088
;  CURRENT APPLICATION NUMBER: US/09/732,091
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 44
;   LENGTH: 265
;   TYPE: PRT
;   ORGANISM: Helicobacter sp.
US-09-732-091-44

  Query Match             100.0%;  Score 1279;  DB 9;  Length 265;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-115;
  Matches  253;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 72

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 132

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 192

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 252

Qy        241 ANGDKKSLQIESI 253
              |||||||||||||
Db        253 ANGDKKSLQIESI 265


RESULT 3
US-10-335-977-9162
; Sequence 9162, Application US/10335977
; Publication No. US20040052799A1
;   GENERAL INFORMATION:
;        APPLICANT: DOUGLAS SMITH et al
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
;                            RELATING TO HELICOBACTER PYLORI FOR
;                            DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 10031
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: LAHIVE & COCKFIELD
;             STREET: 28 State Street
;             CITY: Boston
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02109-1875
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: IBM PC Compatible
;             OPERATING SYSTEM: Windows NT 4.0
;             SOFTWARE: UNIX
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/335,977
;             FILING DATE: 30-Dec-2002
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/993,002
;             FILING DATE: 17-DEC-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mandragouras, Amy E.
;             REGISTRATION NUMBER: 36,207
;             REFERENCE/DOCKET NUMBER: GTN-018
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617)227-7400
;             TELEFAX: (617)742-4214
;   INFORMATION FOR SEQ ID NO: 9162:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 253 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Helicobacter pylori
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...253
;        SEQUENCE DESCRIPTION: SEQ ID NO: 9162:
US-10-335-977-9162

  Query Match             99.3%;  Score 1270;  DB 12;  Length 253;
  Best Local Similarity   99.2%;  Pred. No. 3.6e-114;
  Matches  251;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              || |||||||||:
Db        241 ANEDKKSLQIESV 253


RESULT 4
US-10-335-977-9163
; Sequence 9163, Application US/10335977
; Publication No. US20040052799A1
;   GENERAL INFORMATION:
;        APPLICANT: DOUGLAS SMITH et al
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
;                            RELATING TO HELICOBACTER PYLORI FOR
;                            DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 10031
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: LAHIVE & COCKFIELD
;             STREET: 28 State Street
;             CITY: Boston
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02109-1875
```

```
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: CD/ROM ISO9660
;              COMPUTER: IBM PC Compatible
;              OPERATING SYSTEM: Windows NT 4.0
;              SOFTWARE: UNIX
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/10/335,977
;              FILING DATE: 30-Dec-2002
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/993,002
;              FILING DATE: 17-DEC-1997
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Mandragouras, Amy E.
;              REGISTRATION NUMBER: 36,207
;              REFERENCE/DOCKET NUMBER: GTN-018
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (617)227-7400
;              TELEFAX: (617)742-4214
;    INFORMATION FOR SEQ ID NO: 9163:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 253 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         HYPOTHETICAL: YES
;         ORIGINAL SOURCE:
;              ORGANISM: Helicobacter pylori
;         FEATURE:
;              NAME/KEY:  misc_feature
;              LOCATION: (B) LOCATION 1...253
;         SEQUENCE DESCRIPTION: SEQ ID NO: 9163:
US-10-335-977-9163

  Query Match             99.3%;  Score 1270;  DB 12;  Length 253;
  Best Local Similarity   99.2%;  Pred. No. 3.6e-114;
  Matches  251;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              || |||||||||:
Db        241 ANEDKKSLQIESV 253


RESULT 5
US-10-335-977-9164
; Sequence 9164, Application US/10335977
; Publication No. US20040052799A1
;   GENERAL INFORMATION:
;         APPLICANT: DOUGLAS SMITH et al
;         TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
;                             RELATING TO HELICOBACTER PYLORI FOR
;                             DIAGNOSTICS AND THERAPEUTICS
;         NUMBER OF SEQUENCES: 10031
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: LAHIVE & COCKFIELD
;              STREET: 28 State Street
;              CITY: Boston
;              STATE: Massachusetts
;              COUNTRY: USA
;              ZIP: 02109-1875
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: CD/ROM ISO9660
;              COMPUTER: IBM PC Compatible
;              OPERATING SYSTEM: Windows NT 4.0
;              SOFTWARE: UNIX
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/10/335,977
;              FILING DATE: 30-Dec-2002
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/993,002
;              FILING DATE: 17-DEC-1997
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Mandragouras, Amy E.
;              REGISTRATION NUMBER: 36,207
;              REFERENCE/DOCKET NUMBER: GTN-018
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (617)227-7400
;              TELEFAX: (617)742-4214
;    INFORMATION FOR SEQ ID NO: 9164:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 256 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         HYPOTHETICAL: YES
;         ORIGINAL SOURCE:
;              ORGANISM: Helicobacter pylori
;         FEATURE:
;              NAME/KEY:  misc_feature
;              LOCATION: (B) LOCATION 1...256
;         SEQUENCE DESCRIPTION: SEQ ID NO: 9164:
US-10-335-977-9164

  Query Match             99.3%;  Score 1270;  DB 12;  Length 256;
  Best Local Similarity   99.2%;  Pred. No. 3.6e-114;
  Matches  251;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 63

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 123

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 183

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 243

Qy        241 ANGDKKSLQIESI 253
              || |||||||||:
Db        244 ANEDKKSLQIESV 256


RESULT 6
US-10-335-977-7699
; Sequence 7699, Application US/10335977
; Publication No. US20040052799A1
;   GENERAL INFORMATION:
;         APPLICANT: DOUGLAS SMITH et al
;         TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
;                             RELATING TO HELICOBACTER PYLORI FOR
;                             DIAGNOSTICS AND THERAPEUTICS
;         NUMBER OF SEQUENCES: 10031
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: LAHIVE & COCKFIELD
;              STREET: 28 State Street
```

; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7699:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...248
; SEQUENCE DESCRIPTION: SEQ ID NO: 7699:
US-10-335-977-7699

Query Match 56.5%; Score 722; DB 12; Length 248;
Best Local Similarity 60.3%; Pred. No. 2.8e-61;
Matches 141; Conservative 36; Mismatches 53; Indels 4; Gaps 2;

```
Qy          8 DLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEELQY 67
              |||||:| |||| |||: ||: :||  | ||:||| ||:|:| |||||  |||||||
Db         11 DLEFLKRLSSSDLKDLFDALVYDEDGTLRMNEELTSLTEYQRYGHDYAKYPRRIAEELQR 70

Qy         68 YGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSLEEMDDE 127
              || ||||:| : ||||||||||| || | :|||::: |:|||||||||:|: |||:|
Db         71 YGGNSFANFFRDEGVLYKEILCDACDHLDINYNERSATSLIEQNMLSKLLKDSLEKMSGR 130

Qy        128 EVKEMCDELSIKNTDNL---NRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGL 184
              |:||:|| | : | :   |:| | |: ||||| ||  || ||| ||:|: :  || ||
Db        131 EIKELCDGLGMPNIDKVIGENKQVLIASVLTLFKAGGSHSYALAVAVADAMVRQTLGHGL 190

Qy        185 -SLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLK 237
               |: |   | :||  | ||:||:|||   :|::|||||||:|||::|||| |
Db        191 SSVVGKVALKKTLDILAGPIGWVITGALVSINLAGPAYRVTVPACVLVATLRKK 244
```

RESULT 7
US-10-335-977-7698
; Sequence 7698, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7698:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...237
; SEQUENCE DESCRIPTION: SEQ ID NO: 7698:
US-10-335-977-7698

Query Match 56.0%; Score 716; DB 12; Length 237;
Best Local Similarity 60.1%; Pred. No. 1e-60;
Matches 140; Conservative 36; Mismatches 53; Indels 4; Gaps 2;

```
Qy          9 LEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEELQYY 68
              |||||:| |||| |||: ||: :||  | ||:||| ||:|:| |||||  |||||||
Db          1 LEFLKRLSSSDLKDLFDALVYDEDGTLRMNEELTSLTEYQRYGHDYAKYPRRIAEELQRY 60

Qy         69 GSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSLEEMDDEE 128
              | ||||:| : ||||||||||| || | :|||::: |:|||||||||:|: |||:|
Db         61 GGNSFANFFRDEGVLYKEILCDACDHLDINYNERSATSLIEQNMLSKLLKDSLEKMSGRE 120

Qy        129 VKEMCDELSIKNTDNL---NRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGL- 184
              :||:|| | : | | :   |:| | |: ||||| ||  || ||| ||:|: :  || ||
Db        121 IKELCDGLGMPNIDKVIGENKQVLIASVLTLFKAGGSHSYALAVAVADAMVRQTLGHGLS 180

Qy        185 SLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLK 237
              |: |   | :||  | ||:||:|||   :|::|||||||:|||::|||| |
Db        181 SVVGKVALKKTLDILAGPIGWVITGALVSINLAGPAYRVTVPACVLVATLRKK 233
```

RESULT 8
US-09-882-227-414
; Sequence 414, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Ooomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in t
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615

```
;  PRIOR FILING DATE: 1997-07-29
;  NUMBER OF SEQ ID NOS: 638
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 414
;    LENGTH: 155
;    TYPE: PRT
;    ORGANISM: Helicobacter pylori
US-09-882-227-414

  Query Match             36.9%;  Score 472.5;  DB 10;  Length 155;
  Best Local Similarity   63.3%;  Pred. No. 1.8e-37;
  Matches   93;  Conservative   22;  Mismatches   29;  Indels    3;  Gaps    1;

Qy         38 NEKLTSSIEYKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKV 97
              || ||:| ||||:| |||||  |||||||:|| |||:| : |||||||||| || |||
Db          2 NEDLTNSTEYKRYGHDYAKYPRRIAEELQHYGGNSFANFFRDEGVLYKEILCDACDHLKV 61

Qy         98 NYNKKTETTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNL---NRQALSAATL 154
              |||::: |:|||||||||:|: |||:|   |:||:|:|| : | | :   |:| | |:||
Db         62 NYNEESATSLIEQNMLSKLLKDSLEKMSRREIKELCNELGMTNIDKVIGENKQVLIASTL 121

Qy        155 TLFKMGGFKSYQLAVIVANAVAKTILG 181
              |||| ||  || ||| ||:|: :  ||
Db        122 TLFKAGGSHSYALAVSVADAMVRQTLG 148


RESULT 9
US-09-732-091-20
; Sequence 20, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
;  APPLICANT:  Tian, Jing-Hui
;  APPLICANT:  Walker, Richard I.
;  APPLICANT:  Jackson, W. James
;  TITLE OF INVENTION:  Helicobacter proteins, gene sequences and uses
;  TITLE OF INVENTION:  thereof
;  FILE REFERENCE:  7969-088
;  CURRENT APPLICATION NUMBER: US/09/732,091
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 20
;    LENGTH: 49
;    TYPE: PRT
;    ORGANISM: Helicobacter sp.
US-09-732-091-20

  Query Match             19.7%;  Score 252;  DB 9;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-17;
  Matches   49;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        203 VGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQANGDKKSLQIE 251
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQANGDKKSLQIE 49


RESULT 10
US-09-732-091-17
; Sequence 17, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
;  APPLICANT:  Tian, Jing-Hui
;  APPLICANT:  Walker, Richard I.
;  APPLICANT:  Jackson, W. James
;  TITLE OF INVENTION:  Helicobacter proteins, gene sequences and uses
;  TITLE OF INVENTION:  thereof
;  FILE REFERENCE:  7969-088
;  CURRENT APPLICATION NUMBER: US/09/732,091
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 17
;    LENGTH: 38
;    TYPE: PRT
;    ORGANISM: Helicobacter sp.
US-09-732-091-17

  Query Match             15.5%;  Score 198;  DB 9;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-12;
  Matches   38;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         53 DYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCD 90
              ||||||||||||||||||||||||||||||||||||||
Db          1 DYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCD 38


RESULT 11
US-09-732-091-19
; Sequence 19, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
;  APPLICANT:  Tian, Jing-Hui
;  APPLICANT:  Walker, Richard I.
;  APPLICANT:  Jackson, W. James
;  TITLE OF INVENTION:  Helicobacter proteins, gene sequences and uses
;  TITLE OF INVENTION:  thereof
;  FILE REFERENCE:  7969-088
;  CURRENT APPLICATION NUMBER: US/09/732,091
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 19
;    LENGTH: 41
;    TYPE: PRT
;    ORGANISM: Helicobacter sp.
US-09-732-091-19

  Query Match             15.2%;  Score 194;  DB 9;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-11;
  Matches   41;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        145 NRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGLS 185
              |||||||||||||||||||||||||||||||||||||||||||
Db          1 NRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGLS 41


RESULT 12
US-09-732-091-18
; Sequence 18, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
;  APPLICANT:  Tian, Jing-Hui
;  APPLICANT:  Walker, Richard I.
;  APPLICANT:  Jackson, W. James
;  TITLE OF INVENTION:  Helicobacter proteins, gene sequences and uses
;  TITLE OF INVENTION:  thereof
;  FILE REFERENCE:  7969-088
;  CURRENT APPLICATION NUMBER: US/09/732,091
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 18
;    LENGTH: 30
;    TYPE: PRT
;    ORGANISM: Helicobacter sp.
US-09-732-091-18

  Query Match             11.9%;  Score 152;  DB 9;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-07;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALS 150
              ||||||||||||||||||||||||||||||
Db          1 LEEMDDEEVKEMCDELSIKNTDNLNRQALS 30
```

```
RESULT 13
US-09-732-091-16
; Sequence 16, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
;  APPLICANT:  Tian, Jing-Hui
;  APPLICANT:  Walker, Richard I.
;  APPLICANT:  Jackson, W. James
;  TITLE OF INVENTION:  Helicobacter proteins, gene sequences and uses
;  TITLE OF INVENTION:  thereof
;  FILE REFERENCE:  7969-088
;  CURRENT APPLICATION NUMBER: US/09/732,091
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 16
;   LENGTH: 30
;   TYPE: PRT
;   ORGANISM: Helicobacter sp.
US-09-732-091-16

  Query Match             11.7%;  Score 150;  DB 9;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-07;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFG 30
              ||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFG 30


RESULT 14
US-10-282-122A-43832
; Sequence 43832, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43832
;   LENGTH: 1009
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-10-282-122A-43832

  Query Match             7.7%;  Score 98.5;  DB 12;  Length 1009;
  Best Local Similarity   22.5%;  Pred. No. 3.2;
  Matches   66;  Conservative   45;  Mismatches   83;  Indels   99;  Gaps   16;

Qy          5 YDRDLEF--------LKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAK 56
              |: |||          : :|  :|  |: |:: :  : |:          :| ||  | | |
Db        708 YNNDLELNEQSIEMEMSRLNLTDDNDINEIIAWRGEQEE---------LEQKR--DTYKK 756

Qy         57 YAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTE--TTLIE----- 109
                 |:             |  |      ||:|    ||||  | :| |   |||:
Db        757 RYHEFEMEIA-----RLESLTKD-----KELLDS--DKLKDEYEQKKEKMNTLIDEYSAV 804

Qy        110 ----------QNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLN----------RQA 148
                       |:::| |   : |  | :|: :: : :| ||  ||           |
Db        805 HYQCQNNINKTQSIVSHINYLNQELKDQQEIFQLAEIVSGKNNKNLTLENFVLIYYLDQI 864

Qy        149 LSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGLS--------LAGNQVLTRTLSFLT 200
              :: | | |  |      |||       : :  :  |||        |  |:  :| :| |:
Db        865 IAQANLRLATMSD-NRYQL-------IRREAVSHGLSGLEIDVFDLHSNK--SRHISSLS 914

Qy        201 GPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQANGDKKSLQIESI 253
              |                |  :: ::   : :: :    || :|    : :|||
Db        915 G----------------GETFQSSLALALGLSEI---VQQQSG---GISLESI 945


RESULT 15
US-10-437-963-191043
; Sequence 191043, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated Wit
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191043
;   LENGTH: 815
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_8739C.1.pep
US-10-437-963-191043

  Query Match             7.6%;  Score 97.5;  DB 16;  Length 815;
  Best Local Similarity   28.1%;  Pred. No. 2.9;
  Matches   45;  Conservative   26;  Mismatches   62;  Indels   27;  Gaps    8;

Qy          2 AYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAERI 61
              | : ::||     |: | | | |  :  |   :    | :: ::      |:     | :
Db        531 ANRMNQDL----VLQISKLQDEFNEMQEGLTNKLGEVESVSKAL-----SDELVSVKEMV 581

Qy         62 ---AEELQYYGSNSPASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILE 118
                 |||:   ||  || ::    | ||:| ||  ||      : |   |:::       |
Db        582 HKGQEELE-ATSNELASIVEARDNLKKELL-DVFKKL-----ESTSQELVDERKTVTTLN 634

Qy        119 RSLE------EMDDEEVKEMCDEL--SIKNTDNLNRQALS 150
              | ||      :|| |  | :  :|  : |: | :|| |||
```

Db 635 RELEALVKQLQMDSEARKALEADLDEATKSLDEMNRSALS 674

Search completed: July 5, 2004, 03:13:58
Job time : 34 secs

```
Db          2 NEDLTNSTEYKRYGHDYAKYPRRIAEELQHYGGNSFANFFRDEGVLYKEILCDACDHLKV 61

Qy         98 NYNKKTETTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNL---NRQALSAATL 154

Db         62 NYNEESATSLIEQNMLSKLLKDSLEKMSRREIKELCNELGMTNIDKVIGENKQVLIASTL 121

Qy        155 TLFKMGGFKSYQLAVIVANAVAKTILG 181

Db        122 TLFKAGGSHSYALAVSVADAMVRQTLG 148


RESULT 6
C64721
hypothetical protein b0011 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: C64721; D56688; S28462
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: C64721
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-237 <BLAT>
A;Cross-references: GB:AE000112; GB:U00096; NID:g1786192; PIDN:AAC73122.1; PID:g1786193;
A;Experimental source: strain K-12, substrain MG1655
R;James, R.; Dean, D.O.; Debbage, J.
DNA Seq. 3, 327-332, 1993
A;Title: Five open reading frames upstream of the dnaK gene of Escherichia coli.
A;Reference number: A56688; MUID:94003405; PMID:8400364
A;Accession: D56688
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-189,'F',191-237 <JAM>
A;Cross-references: GB:X67700; NID:g41754; PIDN:CAA47934.1; PID:g41759

  Query Match             20.7%;  Score 265;  DB 2;  Length 237;
  Best Local Similarity   31.2%;  Pred. No. 1.4e-12;
  Matches   77;  Conservative   44;  Mismatches   90;  Indels   36;  Gaps    6;

Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKR---HGDDYAKYAE 59

Db          5 YLNDSDLDFLQHCSEEQLANFARLLTHNEKGKTRLSSVLMRNELFKSMEGHPEQHRRNWQ 64

Qy         60 RIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119

Db         65 LIAGELQHFGGDSIANKLRGHGKLYRAILLDVSKRLKLKADKEMSTFEIEQQLLEQFLRN 124

Qy        120 SLEEMDDE-----------EVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLA 168

Db        125 TWKKMDEEHKQEFLHAVDARVNELEELLPLLMKDKLLAKGVS----------HLLSSQLT 174

Qy        169 VIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPAC 228

Db        175 RILRTHAAMSVLGHGL-LRG--------AGLGGPVGAALNGVKA---VSGSAYRVTIPAV 222

Qy        229 IVVATLR 235

Db        223 LQIACLR 229


RESULT 7
AC0503
conserved hypothetical protein STY0010 [imported] - Salmonella enterica subsp. enterica
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0503
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0503
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-237 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD01163.1; PID:g16501293; GSPDB:GN00176
C;Genetics:
A;Gene: STY0010

  Query Match             20.7%;  Score 265;  DB 2;  Length 237;
  Best Local Similarity   32.6%;  Pred. No. 1.4e-12;
  Matches   79;  Conservative   40;  Mismatches  101;  Indels   22;  Gaps    7;

Qy          1 MAYKYDRDLEPLKQLESSDLLDLFEVLVFGKDGEKR------HNEKLTSSIEYKRHGDDY 54

Db          3 VTYLHDEDLDFLQHCSEEQLADFARLLTHNEKGKARLSSVLSHNE-LFKAME--GHPEQH 59

Qy         55 AKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLS 114

Db         60 RRNWQLIAGEFQHYGGDSIANKLRGHGKQYRAILLDVAKRLKLKADKSMSTFEIEQQLLE 119

Qy        115 KILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFK-MGGFKSYQLAVIVAN 173

Db        120 HFLRHTWQKMDAAHKQEFLQAVDAKVSELEELLPLLMKDRSLAKGVSHLLSTQLTRILRT 179

Qy        174 AVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVAT 233

Db        180 HAAMSILGHGL-LRG--------AGLGGPVGAALNGVKA---MSGSAYRVTIPAVLQIAC 227

Qy        234 LR 235

Db        228 LR 229


RESULT 8
C90630
probable oxidoreductase ECs0012 [imported] - Escherichia coli (strain O157:H7, substr
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: C90630
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90630
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-237 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33434.1; PID:g13359467; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0012

  Query Match             20.5%;  Score 262;  DB 2;  Length 237;
  Best Local Similarity   30.8%;  Pred. No. 2.3e-12;
  Matches   76;  Conservative   45;  Mismatches   90;  Indels   36;  Gaps    6;

Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKR---HGDDYAKYAE 59

Db          5 YLNDSDLDFLQHCSEEQLANFARLLTHNEKGKTRLSSVLMRNELFKSMEGHPEQHRRNWQ 64

Qy         60 RIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119

Db         65 LIAGELQHFGGDSIANKLRGHGKLYRAILLDVSKRLKLKADKEMSTFEIEQQLLEQFLRN 124

Qy        120 SLEEMDDE-----------EVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLA 168

Db        125 TWKKMDEEHKQEFLHAVDARVNELEELLPLLMKDKLLAKGVS----------HLLSSQLT 174
```

```
Qy        169 VIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPAC 228
              |:    | ::|| || | |        : | |||  : ||     ::| :|||||||
Db        175 RILRTHAAMSVLGHGL-LRG--------AGLGGPVGAALNGVKA---VSGSSYRVTIPAV 222

Qy        229 IVVATLR 235
              : :| ||
Db        223 LQIACLR 229
```

RESULT 9
C85481
probable oxidoreductase Z0011 [imported] - Escherichia coli (strain O157:H7, substrain E
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: C85481
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85481
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-237 <STO>
A;Cross-references: GB:AE005174; NID:g12512689; PIDN:AAG54311.1; GSPDB:GN00145; UWGP:Z00
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0011

  Query Match 20.5%; Score 262; DB 2; Length 237;
  Best Local Similarity 30.8%; Pred. No. 2.3e-12;
  Matches 76; Conservative 45; Mismatches 90; Indels 36; Gaps 6;

```
Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKR---HGDDYAKYAE 59
              |  | ||:||:       | :   :|   : |: | :  |  :  :|    | : : :  :
Db          5 YLNDSDLDFLQHCSEEQLANFARLLTHNEKGKTRLSSVLMRNELFKSMEGHPEQHRRNWQ 64

Qy         60 RIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119
               || |||::| :| |: ::| | ||: || ||  :||:  :|:  |  ||| :| : |
Db         65 LIAGELQHFGGDSIANKLRGHGKLYRAILLDVSKRLKLKADKEMSTFEIEQQLLEQFLRN 124

Qy        120 SLEEMDDE-----------EVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLA 168
              : ::|:|:|           | |: : | :   | |  : :|                | ||
Db        125 TWKKMDEEHKQEFLHAVDARVNELEELLPLLMKDKLLAKGVS----------HLLSSQLT 174

Qy        169 VIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPAC 228
              |:    | ::|| || | |        : | |||  : ||     ::| :|||||||
Db        175 RILRTHAAMSVLGHGL-LRG--------AGLGGPVGAALNGVKA---VSGSSYRVTIPAV 222

Qy        229 IVVATLR 235
              : :| ||
Db        223 LQIACLR 229
```

RESULT 10
F64718
hypothetical protein HP1590 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: F64718
R;Tomb, J.F.; White, C.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: F64718
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-39 <TOM>
A;Cross-references: GB:AE000656; GB:AE000511; NID:g2314771; PIDN:AAD08629.1; PID:g231477

  Query Match 10.4%; Score 133; DB 2; Length 39;
  Best Local Similarity 66.7%; Pred. No. 0.00096;
  Matches 26; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNE 39
              |||:|| ||||||:| |||| |||: ||: :||  | ||
Db          1 MAYRYDSDLEFLKRLSSSDLKDLFDALVYDEDGTLRMNE 39
```

RESULT 11
C89910
hypothetical protein SA1181 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: C89910
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: C89910
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1009 <KUR>
A;Cross-references: GB:BA000018; PID:g13701144; PIDN:BAB42439.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA1181

  Query Match 7.7%; Score 98.5; DB 2; Length 1009;
  Best Local Similarity 22.5%; Pred. No. 23;
  Matches 66; Conservative 45; Mismatches 83; Indels 99; Gaps 16;

```
Qy          5 YDRDLEF--------LKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAK 56
              |: |||          : :|  :|  |: |:: :  : |:           :| ||  | | |
Db        708 YNNDLELNEQSIEMEMSRLNLTDDNDINEIIAWRGEQEE---------LEQKR--DTYKK 756

Qy         57 YAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTE--TTLIE----- 109
                 |:       |  |      ||:|    ||||  | :| |  |||:
Db        757 RYHEFEMEIA-----RLESLTKD-----KELLDS--DKLKDEYEQKKEKMNTLIDEYSAV 804

Qy        110 ----------QNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLN----------RQA 148
                       |:::| |   : |  | :|: :: : :| ||  ||             |
Db        805 HYQCQNNINKTQSIVSHINYLNQELKDQQEIFQLAEIVSGKNNKNLTLENFVLIYYLDQI 864

Qy        149 LSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGLS--------LAGNQVLTRTLSFLT 200
              :: | | |     |||       : :  :  |||         |  |:  :| :| |:
Db        865 IAQANLRLATMSD-NRYQL-------IRREAVSHGLSGLEIDVFDLHSNK--SRHISSLS 914

Qy        201 GPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQANGDKKSLQIESI 253
              |              |  :: ::   : :: :    || :|    : :|||
Db        915 G----------------GETFQSSLALALGLSEI---VQQQSG---GISLESI 945
```

RESULT 12
AE0360
hypothetical protein YPO2963 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AE0360
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AE0360
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-282 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92208.1; PID:g15980920; GSPDB:GN00175

C;Genetics:
A;Gene: YP02963

Query Match 7.7%; Score 98; DB 2; Length 282;
Best Local Similarity 24.7%; Pred. No. 4.9;
Matches 58; Conservative 34; Mismatches 87; Indels 56; Gaps 13;

```
Qy          19 DLLDLFEVLVFGKDGEK-RHNEKLTSSIEYKRHGD-DYAKYAERIAEELQYYGSNSFAS- 75
Db           7 EVLDAVERNPYSNSEDKTRQRNNVMRSIMYNKQNELDQVK--------------NSIASK 52

Qy          76 FIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSL-------EEMDDEE 128
Db          53 FIKWS--LIKDQFISEVTRYSSDITAKVNTGEITTETAIKLLDKELSDLRNQDEEMIRER 110

Qy         129 VKE-MCDELSIKNTDNLNRQALSAATLTLFKMG-GFKSYQLAVI----VANAVAKTILGR 182
Db         111 VKQAVIIKASVKENNNVERNEV----INLVFAGVGFVTAGLQIVAGVGMVGSVVGSIPGT 166

Qy         183 GLSLAG-NQVLTRTLSFL-----TGPVGWIITGV--------------WTAIDIA 217
Db         167 LLILHGANNIVESGYYILYRESYTGPVKFVYEGVGAQFGLSKSDSAIMYTAVDVA 221
```

RESULT 13
G90546
conserved hypothetical protein MYPJ_2790 [imported] - Mycoplasma pulmonis (strain UAB CT
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: G90546
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: G90546
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1099 <KUR>
A;Cross-references: GB:AL445566; PID:g14089692; PIDN:CAC13452.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_2790
A;Genetic code: SGC3

Query Match 7.7%; Score 98; DB 2; Length 1099;
Best Local Similarity 22.2%; Pred. No. 28;
Matches 53; Conservative 40; Mismatches 90; Indels 56; Gaps 9;

```
Qy           5 YDRDLEF----LKQLESSDLLDLFEVLVFGKDGEKRHNE----------------KLTS 43
Db          54 FHNDLKIFEGSLKNIAGKDGQNLYEIVIDKFQMKKHKNEIVEKETNEVYNEFFIDPKTKK 113

Qy          44 SIEYKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDV-----CDKLKVN 98
Db         114 RIEYKKR----KKWIGTKEYILEINNSDPEKEYFQPQKTSYDSGLVDIPLFYYNNKNSQT 169

Qy          99 YNKKT--------ETTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALS 150
Db         170 CNQKTTQVNKSSQEKTNKELNMKSSSITKEINQSAESKTTSQNQDLSLNEQNNLNQSSVT 229

Qy         151 A-----ATLTLF-KMGGFKSYQLAVIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPV 203
Db         230 SQNTFEATSSSFNQMPNFDTSQTS-------AQTI-----SPAPNQNLSQTNEVTPSPV 276
```

RESULT 14
C81380
probable D-2-hydroxyacid dehydrogenase Cj0373 [imported] - Campylobacter jejuni (strain
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: C81380
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: C81380
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-311 <PAR>
A;Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74209.1; PID:g696
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0373
C;Superfamily: phosphoglycerate dehydrogenase

Query Match 7.6%; Score 97.5; DB 2; Length 311;
Best Local Similarity 23.0%; Pred. No. 6;
Matches 37; Conservative 34; Mismatches 37; Indels 53; Gaps 8;

```
Qy          40 KLTSSIEYKRH--------GDDYAKYAERIAEELQYY---GSNSFASFIKGEGVLYKEIL 88
Db         138 RILNTLSGKKHGIIGLGTIGKEVAKISKAFGAEIYYYSTSGANKNADFVHLE---LKDLL 194

Qy          89 CDVCDKLKVN--YNKKTETTL----------------------IEQNMLSKILER----- 119
Db         195 -KTCDIISIHAPLNEKTKNLLAFEELKLLKDNAILINVGRGGIVNENDLAKIIDEKNIRV 253

Qy         120 SLEEMDDEEVKEMCDELSIKNTDNL---------NRQALSA 151
Db         254 GLDVLEIEPMMKNHPLLSIKNKENLIITPHVAWASKEALNA 294
```

RESULT 15
A81385
probable ATP /GTP binding protein Cj0411 [imported] - Campylobacter jejuni (strain NC
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: A81385
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Bar
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: A81385
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-728 <PAR>
A;Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74247.1; PID:g696
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0411

Query Match 7.6%; Score 97.5; DB 2; Length 728;
Best Local Similarity 25.4%; Pred. No. 18;
Matches 43; Conservative 27; Mismatches 56; Indels 43; Gaps 8;

```
Qy          10 EFLKQLESSDLL---DLFEVLVFGKDGEKRH----NEKLTSSIEY-----KRHGDDYAKY 57
Db         352 KFLIVLTKADLLSKKDLEEVIVYTKESLKSRLVDLDENLVEKIDFLCVSAKMASDFYKGL 411

Qy          58 AERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKK-----TETTLIEQNM 112
Db         412 ASK--ESLQKSGMQEF------ENYLFNELYAGEKSKIALRAYKKELHLELKNILSEYEM 463

Qy         113 LSKILERSLEEMDDEE-------------VKEMCDELS-----IKNTDN 143
Db         464 QNRLIKENKQGVSEENQKLLLELQKQNTLLKEAQDEISNSIAKLKNIDS 512
```

Search completed: July 5, 2004, 03:13:15
Job time : 14 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 5, 2004, 03:08:23 ; Search time 12 Seconds
(without alignments)
1097.812 Million cell updates/sec

Title: US-09-732-091-4
Perfect score: 1279
Sequence: 1 MAYKYDRDLEFLKQLESSDL.....LRLKTQQANGDKKSLQIESI 253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1279	100.0	253	1	YF88_HELPY	O26107 helicobacte
2	1270	99.3	253	1	YF88_HELPJ	Q9zj24 helicobacte
3	637	49.8	209	1	YF87_HELPY	O26106 helicobacte
4	630	49.3	209	1	YF87_HELPJ	Q9zj25 helicobacte
5	265	20.7	237	1	YAAW_ECOLI	P75617 escherichia
6	262	20.5	237	1	YAAW_ECO57	P58316 escherichia
7	105	8.2	726	1	HS9A_BRARE	Q90474 brachydanio
8	97	7.6	916	1	SECA_TREPA	O83394 treponema p
9	96.5	7.5	373	1	BIOF_AQUAE	O66875 aquifex aeo
10	95.5	7.5	344	1	ABIC_LACLA	Q01457 lactococcus
11	95	7.4	517	1	YC09_CAMJE	Q9pn86 campylobact
12	95	7.4	3214	1	BPA1_HUMAN	Q03001 homo sapien
13	94.5	7.4	1433	1	REST_CHICK	O42184 gallus gall
14	93.5	7.3	802	1	PAC_BACME	Q60136 bacillus me
15	92.5	7.2	802	1	PAC_ARTVI	P31356 arthrobacte
16	92.5	7.2	892	1	RCQ1_CAEEL	P46064 caenorhabdi
17	91.5	7.2	399	1	RPSD_THEMA	P77994 thermotoga
18	91.5	7.2	472	1	GATB_METJA	Q57624 methanococc
19	91.5	7.2	719	1	HS9A_HORSE	Q9gkx7 equus cabal
20	91.5	7.2	1769	1	YJK9_YEAST	P42945 saccharomyc
21	91	7.1	284	1	TPM1_CIOIN	Q07068 ciona intes
22	90	7.0	404	1	LA_BOVIN	P10881 bos taurus
23	89.5	7.0	280	1	Y705_CHLPN	Q9z7k0 chlamydia p
24	89.5	7.0	633	1	GATE_SULSO	Q97zh6 sulfolobus
25	89.5	7.0	869	1	AMPN_HAEIN	P45274 haemophilus
26	89	7.0	386	1	NESG_HUMAN	Q9ul16 homo sapien
27	88.5	6.9	732	1	HS9A_CRIGR	P46633 cricetulus
28	88	6.9	371	1	RPSD_STRMU	O33662 streptococc
29	88	6.9	670	1	SSM4_SCHPO	O42667 schizosacch
30	87.5	6.8	728	1	HS9A_CHICK	P11501 gallus gall
31	87.5	6.8	731	1	HS9A_HUMAN	P07900 homo sapien
32	87.5	6.8	732	1	HS9A_MOUSE	P07901 mus musculu
33	87.5	6.8	732	1	HS9A_PIG	O02705 sus scrofa
34	87	6.8	607	1	HTPG_FUSNN	Q8rgh4 fusobacteri
35	86.5	6.8	619	1	PRIA_HELPJ	Q9zke4 helicobacte
36	86.5	6.8	686	1	HMCT_HELPJ	Q9zl53 helicobacte
37	86.5	6.8	1076	1	CARB_ARCFU	O28994 archaeoglob
38	86.5	6.8	1102	1	CARB_STRCO	Q9kxr6 streptomyce
39	86	6.7	619	1	PRIA_HELPY	O25149 helicobacte
40	86	6.7	633	1	YN40_ARCFU	O27944 archaeoglob
41	86	6.7	1164	1	KEL1_YEAST	P38853 saccharomyc
42	86	6.7	1169	1	EX5B_BORBU	O51578 borrelia bu
43	85.5	6.7	411	1	HFLK_BUCAP	Q8k914 buchnera ap
44	85.5	6.7	1085	1	AMP1_PLAFQ	O96935 plasmodium
45	85	6.6	894	1	SEC3_HUMAN	Q9nv70 homo sapien

ALIGNMENTS

RESULT 1
YF88_HELPY
ID YF88_HELPY STANDARD; PRT; 253 AA.
AC O26107;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein HP1588.
GN HP1588.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
CC -!- SIMILARITY: Belongs to the UPF0174 family.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; AE000656; AAD08627.1; -.
DR PIR; D64718; D64718.
DR TIGR; HP1588; -.
DR InterPro; IPR005367; UPF0174.
DR Pfam; PF03667; UPF0174; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 28417 MW; 00E15A38B1A2036A CRC64;

Query Match 100.0%; Score 1279; DB 1; Length 253;
Best Local Similarity 100.0%; Pred. No. 9.1e-88;
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
Db 1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy 61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

```
Db          61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy         121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Db         121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy         181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Db         181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy         241 ANGDKKSLQIESI 253

Db         241 ANGDKKSLQIESI 253


RESULT 2
YF88_HELPJ
ID   YF88_HELPJ     STANDARD;      PRT;   253 AA.
AC   Q9ZJ24;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein JHP1494.
GN   JHP1494.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- SIMILARITY: Belongs to the UPF0174 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001571; AAD07073.1; -.
DR   PIR; B71800; B71800.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   253 AA;  28475 MW;  127158B2B1A2036A CRC64;

  Query Match           99.3%;  Score 1270;  DB 1;  Length 253;
  Best Local Similarity   99.2%;  Pred. No. 4.2e-87;
  Matches  251;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy           1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Db           1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy          61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Db          61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy         121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Db         121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy         181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Db         181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy         241 ANGDKKSLQIESI 253

Db         241 ANEDKKSLQIESV 253


RESULT 3
YF87_HELPY
ID   YF87_HELPY     STANDARD;      PRT;   209 AA.
AC   O26106;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein HP1587.
GN   HP1587.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
RN   [2]
RP   CONCEPTUAL TRANSLATION.
RA   Bairoch A.;
RL   Unpublished observations (OCT-2001).
CC   -!- SIMILARITY: Belongs to the UPF0174 family.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC       frameshift in position 160.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000656; AAD08626.1; ALT_FRAME.
DR   TIGR; HP1587; -.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   209 AA;  23069 MW;  F98D3FB8F3F62323 CRC64;

  Query Match           49.8%;  Score 637;  DB 1;  Length 209;
  Best Local Similarity   60.3%;  Pred. No. 2.3e-40;
  Matches  123;  Conservative   33;  Mismatches   44;  Indels    4;  Gaps    2;

Qy          38 NEKLTSSIEYKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKV 97

Db           2 NEDLTNSTEYKRYGHDYAKYPRRIAEELQHYGGNSFANFFRDEGVLYKEILCDACDHLKV 61

Qy          98 NYNKKTETTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNL---NRQALSAATL 154

Db          62 NYNEESATSLIEQNMLSKLLKDSLEKMSRREIKELCNELGMTNIDKVIGENKQVLIASTL 121
```

```
Qy        155 TLFKMGGFKSYQLAVIVANAVAKTILGR-GLSLAGNQVLTRTLSFLTGPVGWIITGVWTA 213
              |||| ||  || ||| ||:|: :  ||     : |   | :||  | ||:||:|||   :
Db        122 TLFKAGGSHSYALAVSVADAMVRQTLGHXACYVVGKVALKKTLGVLAGPIGWVITGALVS 181

Qy        214 IDIAGPAYRVTIPACIVVATLRLK 237
              |::|||||||:|||:::||||||
Db        182 INLAGPAYRVTVPACVLIATLRLK 205


RESULT 4
YF87_HELPJ
ID   YF87_HELPJ     STANDARD;      PRT;   209 AA.
AC   Q9ZJ25;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein JHP1493.
GN   JHP1493.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
CC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- SIMILARITY: Belongs to the UPF0174 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001571; AAD07072.1; -.
DR   PIR; A71800; A71800.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   209 AA;  22907 MW;  C6EC950CDD424CAF CRC64;

  Query Match           49.3%;  Score 630;  DB 1;  Length 209;
  Best Local Similarity  59.8%;  Pred. No. 7.7e-40;
  Matches  122;  Conservative   32;  Mismatches   46;  Indels    4;  Gaps    2;

Qy         38 NEKLTSSIEYKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKV 97
              ||:|||  ||:|:|:| |||||  |||||| || ||||:| : ||||||||||| || | :
Db          2 NEELTSLTEYQRYGHDYAKYPRRIAEELQRYGGNSFANFFRDEGVLYKEILCDACDHLDI 61

Qy         98 NYNKKTETTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNL---NRQALSAATL 154
              |||::: |:|||||||||:|: |||:|   |:||:|| | : | | :   |:| | |: |
Db         62 NYNERSATSLIEQNMLSKLLKDSLEKMSGREIKELCDGLGMPNIDKVIGENKQVLIASVL 121

Qy        155 TLFKMGGFKSYQLAVIVANAVAKTILGRGL-SLAGNQVLTRTLSFLTGPVGWIITGVWTA 213
              |||| ||  || ||| ||:|: :  || ||  |: |   | :||  | ||:||:|||   :
Db        122 TLFKAGGSHSYALAVAVADAMVRQTLGHGLSSVVGKVALKKTLDILAGPIGWVITGALVS 181

Qy        214 IDIAGPAYRVTIPACIVVATLRLK 237
              |::|||||||:|||::|||||
Db        182 INLAGPAYRVTVPACVLVATLRKK 205


RESULT 5
YAAW_ECOLI
ID   YAAW_ECOLI     STANDARD;      PRT;   237 AA.
AC   P75617; Q47290;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein yaaW.
GN   YAAW OR B0011.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94003405; PubMed=8400364;
RA   James R., Dean D.O., Debbage J.;
RT   "Five open reading frames upstream of the dnaK gene of E. coli.";
RL   DNA Seq. 3:327-332(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- SIMILARITY: Belongs to the UPF0174 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X67700; CAA47934.1; -.
DR   EMBL; AE000112; AAC73122.1; -.
DR   PIR; C64721; C64721.
DR   EcoGene; EG14340; yaaW.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT    190    190       L -> F (IN REF. 1).
SQ   SEQUENCE   237 AA;  26665 MW;  A36682B2EA116747 CRC64;

  Query Match           20.7%;  Score 265;  DB 1;  Length 237;
  Best Local Similarity  31.2%;  Pred. No. 9.2e-13;
  Matches   77;  Conservative   44;  Mismatches   90;  Indels   36;  Gaps    6;

Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKR---HGDDYAKYAE 59
              | | ||:||:       | :   :|   : |: | :   |  :  :|    | : : :  :
Db          5 YLNDSDLDFLQHCSEEQLANFARLLTHNEKGKTRLSSVLMRNELFKSMEGHPEQHRRNWQ 64

Qy         60 RIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119
               || |||::| :| |: ::| | ||: || ||  :||:  :|:  |  ||| :| : |
Db         65 LIAGELQHFGGDSIANKLRGHGKLYRAILLDVSKRLKLKADKEMSTFEIEQQLLEQFLRN 124

Qy        120 SLEEMDDE----------EVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLA 168
              : ::||:|           | |: : | :   | |  : :|              | ||
Db        125 TWKKMDEEHKQEFLHAVDARVNELEELLPLLMKDKLLAKGVS----------HLLSSQLT 174

Qy        169 VIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPAC 228
              |:    | ::|| || | |       : | ||||  : ||     ::| ||||||||
Db        175 RILRTHAAMSVLGHGL-LRG--------AGLGGPVGAALNGVKA---VSGSAYRVTIPAV 222

Qy        229 IVVATLR 235
              : :| ||
Db        223 LQIACLR 229


RESULT 6
```

```
YAAW_ECO57
ID   YAAW_ECO57     STANDARD;      PRT;   237 AA.
AC   P58316;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein yaaW.
GN   YAAW OR Z0011 OR ECS0012.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- SIMILARITY: Belongs to the UPF0174 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005178; AAG54311.1; -.
DR   EMBL; AP002550; BAB33434.1; -.
DR   PIR; C85481; C85481.
DR   PIR; C90630; C90630.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   237 AA;  26681 MW;  A25482B2EA116759 CRC64;

  Query Match             20.5%;  Score 262;  DB 1;  Length 237;
  Best Local Similarity   30.8%;  Pred. No. 1.5e-12;
  Matches   76;  Conservative   45;  Mismatches   90;  Indels   36;  Gaps    6;

Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKR---HGDDYAKYAE 59
              |  | |:||:       | :   :|   : |: | :  |  :  :|    | : :  :
Db          5 YLNDSDLDFLQHCSEEQLANFARLLTHNEKGKTRLSSVLMRNELFKSMEGHPEQHRRNWQ 64

Qy         60 RIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119
               || |||::| :  |: ::| | ||: || ||  :||:  :|:  |  ||| :| : |
Db         65 LIAGELQHFGGDSIANKLRGHGKLYRAILLDVSKRLKLKADKEMSTFEIEQQLLEQFLRN 124

Qy        120 SLEEMDDE-----------EVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLA 168
              : ::||:|           | |: : | :   | |  : :|             | ||
Db        125 TWKKMDEEHKQEFLHAVDARVNELEELLPLLMKDKLLAKGVS----------HLLSSQLT 174

Qy        169 VIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPAC 228
               |:    | ::|| || : |        : | |||  : ||     ::| :|||||
Db        175 RILRTHAAMSVLGHGL-LRG--------AGLGGPVGAALNGVKA---VSGSSYRVTIPAV 222

Qy        229 IVVATLR 235
              : :| ||
Db        223 LQIACLR 229


RESULT 7
HS9A_BRARE
ID   HS9A_BRARE     STANDARD;      PRT;   726 AA.
AC   Q90474;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Heat shock protein HSP 90-alpha.
GN   HSP90A OR HSP90.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lele Z., Hadfi S., Sass J.B., Krone P.H.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 36-136 FROM N.A.
RX   MEDLINE=95071389; PubMed=7980538;
RA   Krone P.H., Sass J.B.;
RT   "HSP 90 alpha and HSP 90 beta genes are present in the zebrafish and
RT   are differentially regulated in developing embryos.";
RL   Biochem. Biophys. Res. Commun. 204:746-752(1994).
CC   -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the heat shock protein 90 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF068773; AAC21567.1; -.
DR   EMBL; L35586; AAA97518.1; -.
DR   PIR; JC2343; JC2343.
DR   HSSP; P07900; 1BYQ.
DR   ZFIN; ZDB-GENE-990415-94; hsp90a.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR001404; Hsp90.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00183; HSP90; 1.
DR   PRINTS; PR00775; HEATSHOCK90.
DR   SMART; SM00387; HATPase_c; 1.
DR   PROSITE; PS00298; HSP90; 1.
KW   Chaperone; ATP-binding; Heat shock.
SQ   SEQUENCE   726 AA;  83561 MW;  F7DEB8EF1FBC9CEC CRC64;

  Query Match             8.2%;  Score 105;  DB 1;  Length 726;
  Best Local Similarity   21.4%;  Pred. No. 2.5;
  Matches   61;  Conservative   41;  Mismatches   75;  Indels  108;  Gaps   13;

Qy         21 LDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEELQYYGSNS-------- 72
              ||||  |   ||  |:: |: : :|:   | |  ::  ::::: |:|| | |
Db        415 LDLFTELAEDKDNYKKYYEQFSKNIKLGIHED--SQNRKKLSDLLRYYTSASGDEMVSLK 472

Qy         73 -FAS----------FIKGE-------------------GVLYK-EILCDVCDKLKVNYNK 101
               :|          :| ||                   |:|  | : : |  :    |:
Db        473 DYVSRMKDTQKHIYYITGETKDQVANSAFVERLRKAGLEVIYMIEPIDEYCVQQLKEYDG 532

Qy        102 KTETTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGG 161
              |       |::| : :  ||  :||| |:  |||  |  :|  :
```

```
Db        533 K--------NLVS-VTKEGLELPEDEEEKKKQDELKAK-YENLCK---------------- 567

Qy        162 FKSYQLAVIVANAVAKTILGRGLS--LAGNQVLTRTLSFLTGPVGWIIT------------ 208
                          : | || : :       |::::     :|   ||
Db        568 -------------IMKDILDKKIEKVTVSNRLVSSPCCIVTSTYGWTANMERIMKSQALR 614

Qy        209 -----GVWTA---IDIAGPAYRVTIPACIVVATLRLKTQQANGDK 245
                  |  ||   ::|          ||  :| ||| | :    ||
Db        615 DNSTMGYMTAKKHLEIN--------PAHPIVETLREKAEAEKNDK 651


RESULT 8
SECA_TREPA
ID   SECA_TREPA     STANDARD;      PRT;   916 AA.
AC   O83394;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Preprotein translocase secA subunit.
GN   SECA OR TP0379.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Nichols;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- FUNCTION: Involved in protein export. Interacts with the secY/secE
CC       subunits. SecA has a central role in coupling the hydrolysis of
CC       ATP to the transfer of pre-secretory periplasmic and outer
CC       membrane proteins across the membrane (By similarity).
CC   -!- SUBUNIT: Part of the prokaryotic protein translocation apparatus
CC       which comprise secA, secB, secD, secE, secF, secG and secY (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic side of plasma membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the secA family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001217; AAC65365.1; -.
DR   PIR; E71330; E71330.
DR   TIGR; TP0379; -.
DR   InterPro; IPR004027; SEC_C_motif.
DR   InterPro; IPR000185; SecA.
DR   Pfam; PF02810; SEC-C; 1.
DR   Pfam; PF01043; SecA_protein; 1.
DR   PRINTS; PR00906; SECA.
DR   TIGRFAMs; TIGR00963; secA; 1.
DR   PROSITE; PS01312; SECA; 1.
KW   Protein transport; ATP-binding; Membrane; Translocation; Transport;
KW   Complete proteome.
FT   NP_BIND     103    110       ATP (POTENTIAL).
SQ   SEQUENCE   916 AA;  103779 MW;  EA5561F6EE7C65AE CRC64;

  Query Match             7.6%;  Score 97;  DB 1;  Length 916;
  Best Local Similarity   21.8%;  Pred. No. 13;
  Matches   52;  Conservative   44;  Mismatches   67;  Indels   76;  Gaps   13;

Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGE-------------KRHNEKLTSSIEYK 48
                :||    : :  |    : ||   :|     ||:                 :|:::  |  :|| |
Db        311 FKY---IHYFTQALRAHLLYRADVDYVVKDGQVQIVDEFTGRILEGRRYSDGLHQAIEAK 367

Qy         49 RHGDDYAKYAERIAEELQYYGSNSPASFIKGEGVLYKEI--LCDVCDKLKVNYNK--KTE 104
                |         |||:  :   : :| :|  :    :||::  :    |   :  ||  | |
Db        368 EH--------IRIAQRNRTMATITFQNPFR----MYKKLSGMTGTADTEALELNKIYKLE 415

Qy        105 TTLIEQNM-LSKILERSLEEMDDEEV-KEMCDE---------------LSIKNTDNLNRQ 147
                  ::  |: :::: |  :  : :||    :|||                 :||: :: |:
Db        416 VVVLPTNLPVARVDEHDVVYLSEEEKWSAICDEIKEAHTRGQPVLVGTISIEKSEKLS-- 473

Qy        148 ALSAATLTLFKMGGFKSYQL--------AVIVANA-------VAKTILGRG--LSLAGN 189
                    | :  | |   |          |:|:| |       :|  : |||  : | ||
Db        474 -------ALLRTRGVKHEVLNAKNHAREALIIAEAGAKGSVTIATNMAGRGTDIKLGGN 525


RESULT 9
BIOF_AQUAE
ID   BIOF_AQUAE     STANDARD;      PRT;   373 AA.
AC   O66875;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   8-amino-7-oxononanoate synthase (EC 2.3.1.47) (AONS) (8-amino-7-
DE   ketopelargonate synthase) (7-keto-8-amino-pelargonic acid synthetase)
DE   (7-KAP synthetase) (L-alanine--pimelyl CoA ligase).
GN   BIOF OR AQ_626.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olsen G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- CATALYTIC ACTIVITY: 6-carboxyhexanoyl-CoA + L-alanine = 8-amino-7-
CC       oxononanoate + CoA + CO(2).
CC   -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC   -!- PATHWAY: Biotin biosynthesis; first step.
CC   -!- SIMILARITY: Belongs to class-II of pyridoxal-phosphate-dependent
CC       aminotransferases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000699; AAC06836.1; -.
DR   PIR; G70355; G70355.
DR   HSSP; P12998; 1BS0.
DR   InterPro; IPR003408; Ala_synthase.
DR   InterPro; IPR004839; Aminotrans_I/II.
DR   InterPro; IPR001917; Aminotrans_II.
DR   InterPro; IPR004723; BioF.
DR   Pfam; PF02490; ALA_synthase; 1.
DR   Pfam; PF00155; aminotran_1_2; 1.
DR   TIGRFAMs; TIGR00858; bioF; 1.
DR   PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
KW   Biotin biosynthesis; Transferase; Pyridoxal phosphate;
KW   Complete proteome.
FT   BINDING     223    223       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   373 AA;  42532 MW;  536B34A5D5F84401 CRC64;
```

```
  Query Match             7.5%;  Score 96.5;  DB 1;  Length 373;
  Best Local Similarity   22.6%;  Pred. No. 4.7;
  Matches   64;  Conservative   40;  Mismatches  102;  Indels   77;  Gaps   14;

Qy          9 LEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHG----------DDYAKYA 58
              || :|   |:| |           | ::| | :  ||   :              | |:
Db         23 LEGVKDFCSNDYL----------GLRKHPEVVEESIRVLKEAGLGSGASQLVSGYTKHH 71

Qy         59 ERIAEEL-QYYGSNSFASFIKG------------EGVLYKEILCDVCDKLKVNYNKKTE 104
                : |:| :: |: |   | |              || |   :| |     ::|:
Db         72 RELEEKLAEFKGTESCVLFGSGFLANVGTIPALVEEGDL---VLSD-----ELNH----- 118

Qy        105 TTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGG--- 161
               ::|:   |||  :|  :  | ||::|      ||      |  :   | |:| | |
Db        119 ASIIDGVRLSKAQKRVFKHKDYEELEEFLK----KNRKKFRRVLI--ITDTVFSMDGDVA 172

Qy        162 --------FKSYQLAVIVANAVAKTILGR-GLSLAGNQ-----VLTRTLSFLTGPVGWII 207
                     : |   : :  |      :|: ||   | :     ::  |||   |  |  :
Db        173 DLKRLTQICEEYDCMLYIDEAHTTGTIGKGGLDYFGIEHKEYIIVMGTLSKALGSYGAFV 232

Qy        208 TGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQANGDKKSLQI 250
              |    ||      | |      :: :|  |       | ||:::|
Db        233 CGTKLLID-----YLVNKARSLIFST-SLPPSVCAGAKKAIEI 269


RESULT 10
ABIC_LACLA
ID   ABIC_LACLA     STANDARD;      PRT;   344 AA.
AC   Q01457;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-1993 (Rel. 27, Last annotation update)
DE   Abortive phage resistance protein abiC.
GN   ABIC OR PRF.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG   Plasmid pTN20.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ME2;
RX   MEDLINE=93054365; PubMed=1429469;
RA   Durmaz E., Higgins D.L., Klaenhammer T.R.;
RT   "Molecular characterization of a second abortive phage resistance
RT   gene present in Lactococcus lactis subsp. lactis ME2.";
RL   J. Bacteriol. 174:7463-7469(1992).
CC   -!- FUNCTION: PROVIDES RESISTANCE TO BACTERICPHAGE BY ABORTIVE
CC       INFECTION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M95956; AAA53569.1; -.
DR   PIR; A47025; A47025.
KW   Plasmid.
SQ   SEQUENCE   344 AA;  40128 MW;  AA6E41E71B4A003C CRC64;

  Query Match             7.5%;  Score 95.5;  DB 1;  Length 344;
  Best Local Similarity   23.8%;  Pred. No. 5.1;
  Matches   50;  Conservative   36;  Mismatches   73;  Indels   51;  Gaps    9;

Qy          7 RDLEFLKQLESSD--------LLDLFEVLVFGKDGEKRHNEKLTSSIEYKR-----HGDD 53
              ::  |  |:| :||        |||||:     |:  |     |  | : |||     ||:
Db         92 KNQEESKKLNNSDSANREFYSLLDLFK-----KEQNKSETIKAISFL-YKRAINDKHGNS 144

Qy         54 YAK-----------YAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKK 102
                :           :   : |  |::     :|  ||:  | ||       |:::  :::
Db        145 FINDYNIDIGNGLDFRFNLFESNQWFSITYNSSKYKGKN-LTKE------QKVELVISRQ 197

Qy        103 TETTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGF 162
                :   : :   |  | |: ::    ::  ||   ||    : |  ||
Db        198 FDDVYNKMSSYFKIFHRILKSLNKRFDEKKLDESDYKNYIGILRTQLS------------ 245

Qy        163 KSYQLAVIVANAV-AKTILGRGLSLAGNQV 191
                | :| ||: |::  |  ||| |: | |  :
Db        246 -SEELVVILINSLYVKRGLGLGIELIGTNL 274


RESULT 11
YC09_CAMJE
ID   YC09_CAMJE     STANDARD;      PRT;   517 AA.
AC   Q9PN86;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical UPF0144 protein Cj1209.
GN   CJ1209.
OS   Campylobacter jejuni.
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Campylobacteraceae; Campylobacter.
OX   NCBI_TaxID=197;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11168;
RX   MEDLINE=20150912; PubMed=10688204;
RA   Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA   Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA   Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA   Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA   Whitehead S., Barrell B.G.;
RT   "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT   reveals hypervariable sequences.";
RL   Nature 403:665-668(2000).
CC   -!- SIMILARITY: Belongs to the UPF0144 family.
CC   -!- SIMILARITY: Contains 1 HD domain.
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL139077; CAB73463.1; -.
DR   PIR; C81327; C81327.
DR   HAMAP; MF_00335; -; 1.
DR   InterPro; IPR006674; HD.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR003607; Met_phsphohydro.
DR   InterPro; IPR006675; Unchar_HDIG.
DR   Pfam; PF01966; HD; 1.
DR   Pfam; PF00013; KH; 1.
DR   SMART; SM00471; HDc; 1.
DR   SMART; SM00322; KH; 1.
DR   TIGRFAMs; TIGR00277; HDIG; 1.
KW   Hypothetical protein; Transmembrane; RNA-binding; Complete proteome.
FT   TRANSMEM      1     21       POTENTIAL.
FT   DOMAIN      206    274       KH.
FT   DOMAIN      333    426       HD.
SQ   SEQUENCE   517 AA;  58025 MW;  6A799602ED363A25 CRC64;

  Query Match             7.4%;  Score 95;  DB 1;  Length 517;
  Best Local Similarity   21.2%;  Pred. No. 9.1;
  Matches   54;  Conservative   43;  Mismatches   98;  Indels   60;  Gaps   10;

Qy          2 AYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTS-SIEYKRHGDDYAKYAER 60
              | :|: :|  ||  ::| |     ||       :|:: ||      :: :  || :|  ::
Db         43 AIEYEAEL-ILKDAKNSILNAELEV-------KKKYEEKTHKPQKDFNQKFDDLSKKEQK 94
```

```
Qy          61 IAEELQYYGSNSFASFIKGEGVLYKEILC----------DVCDKLKVNYNKKTETTL-- 107
               : :| :   :             || ||           || ||||  | :| :  |
Db          95 LQQEEEEKLKED------------KEYLCKSQKHIQNLQSDV-DKLKNKYQEKLDDVLKI 140

Qy         108 ------IEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGG 161
                    : :|   :|: : :||   |::  :  :   :  :   |:|         :  |
Db         141 LEHSTGLTQNEAKEIILKKVEENSREQIAHIVRKYEEEAKNEAKRKANFIIAQATSRFAG 200

Qy         162 FKSYQLAVIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAY 221
                 : :  : | |     : || :   | |  :||         :: ||   ||
Db         201 EFAAERLINVINIKNDELKGRIIGKEGRNV--KTLE--------MVLGVDIIID------ 244

Qy         222 RVTIPACIVVATLRL 236
                   |  |:|:   |
Db         245 --DTPGAIIVSCFNL 257
```

RESULT 12
BPA1_HUMAN
ID BPA1_HUMAN STANDARD; PRT; 3214 AA.
AC Q03001; Q12825; Q13266; Q13267; Q13775; Q96J76; Q96QT5; Q9UGD7;
AC Q9UGD8;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bullous pemphigoid antigen 1 isoforms 1/2/3/4/5/8 (230 kDa bullous
DE pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein) (Dystonia
DE musculorum protein) (Fragment).
GN BPAG1 OR DMH OR DT OR KIAA0728.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Keratinocytes;
RX MEDLINE=92011493; PubMed=1717441;
RA Sawamura D., Li K., Chu M.-L., Uitto J.;
RT "Human bullous pemphigoid antigen (BPAG1). Amino acid sequences
RT deduced from cloned cDNAs predict biologically important peptide
RT segments and protein domains.";
RL J. Biol. Chem. 266:17784-17790(1991).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Keratinocytes;
RX MEDLINE=93346806; PubMed=8345227;
RA Elgart G.W., Stanley J.R.;
RT "Cloning of the 5' mRNA for the 230-kD bullous pemphigoid antigen by
RT rapid amplification of cDNA ends.";
RL J. Invest. Dermatol. 101:244-246(1993).
RN [3]
RP SEQUENCE OF 1-645 FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Fetal brain, and Retina;
RX MEDLINE=96121394; PubMed=8575775;
RA Brown A., Dalpe G., Mathieu M., Kothary R.;
RT "Cloning and characterization of the neural isoforms of human
RT dystonin.";
RL Genomics 29:777-780(1995).
RN [4]
RP SEQUENCE OF 321-3214 FROM N.A. (ISOFORMS 1 AND 4).
RA Laird G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE OF 392-492 FROM N.A.
RC TISSUE=Pineal gland;
RA Geerts D., Sonnenberg A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE OF 1082-3214 FROM N.A. (ISOFORM 3).
RC TISSUE=Keratinocytes;
RX MEDLINE=91286285; PubMed=1712022;
RA Tanaka T., Parry D.A.D., Klaus-Kovtun V., Steinert P.M.,
RA Stanley J.R.;
RT "Comparison of molecularly cloned bullous pemphigoid antigen to
RT desmoplakin I confirms that they define a new family of cell adhesion
RT junction plaque proteins.";
RL J. Biol. Chem. 266:12555-12559(1991).
RN [7]
RP SEQUENCE OF 2160-2767 FROM N.A.
RX MEDLINE=91216368; PubMed=2090522;
RA Owaribe K., Kartenbeck J., Stumpp S., Magin T.M., Krieg T.,
RA Diaz L.A., Franke W.W.;
RT "The hemidesmosomal plaque. I. Characterization of a major
RT constituent protein as a differentiation marker for certain forms of
RT epithelia.";
RL Differentiation 45:207-220(1990).
RN [8]
RP SEQUENCE OF 2287-3214 FROM N.A. (ISOFORM 5).
RC TISSUE=Keratinocytes;
RX MEDLINE=89067122; PubMed=2461961;
RA Stanley J.R., Tanaka T., Mueller S., Klaus-Kovtun V., Roop D.;
RT "Isolation of complementary DNA for bullous pemphigoid antigen by use
RT of patients' autoantibodies.";
RL J. Clin. Invest. 82:1864-1870(1988).
RN [9]
RP SEQUENCE OF 2462-3214 FROM N.A. (ISOFORM 8).
RX MEDLINE=94280413; PubMed=8010969;
RA Hopkinson S.B., Jones J.C.;
RT "Identification of a second protein product of the gene encoding a
RT human epidermal autoantigen.";
RL Biochem. J. 300:851-857(1994).
RN [10]
RP DOMAINS.
RX MEDLINE=96199235; PubMed=8621649;
RA Tang H.-Y., Chaffotte A.-F., Thacher S.M.;
RT "Structural analysis of the predicted coiled-coil rod domain of the
RT cytoplasmic bullous pemphigoid antigen (BPAG1). Empirical
RT localization of the N-terminal globular domain-rod boundary.";
RL J. Biol. Chem. 271:9716-9722(1996).
RN [11]
RP SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND ALTERNATIVE SPLICING.
RC TISSUE=Keratinocytes;
RX MEDLINE=21839111; PubMed=11751855;
RA Okumura M., Yamakawa H., Ohara O., Owaribe K.;
RT "Novel alternative splicings of BPAG1 (bullous pemphigoid antigen 1)
RT including the domain structure closely related to MACF (microtubule
RT actin cross-linking factor).";
RL J. Biol. Chem. 277:6682-6687(2002).
CC -!- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
CC intermediate filaments to the inner plaque of hemidesmosomes. The
CC proteins may self-aggregate to form filaments or a two-dimensional
CC mesh.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=10;
CC Comment=Isoforms 1, 2, 5 and 8 are or may be fragments;
CC Name=1;
CC IsoId=Q03001-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q03001-2; Sequence=VSP_005053, VSP_005055;
CC Name=3; Synonyms=1e;
CC IsoId=Q03001-3; Sequence=VSP_005054, VSP_005056, VSP_005057,
CC VSP_005058, VSP_005059, VSP_005060,
CC VSP_005061;
CC Name=4;
CC IsoId=Q03001-4; Sequence=VSP_005054, VSP_005056;
CC Name=5;
CC IsoId=Q03001-5; Sequence=VSP_005064, VSP_005065;
CC Name=6; Synonyms=EA;
CC IsoId=O94833-2; Sequence=External;
CC Name=7; Synonyms=EB;
CC IsoId=Q8WXK8-1; Sequence=External;
CC Name=8;
CC IsoId=Q03001-6; Sequence=VSP_005062, VSP_005063;

```
CC       Name=9;
CC         IsoId=O94833-3; Sequence=External;
CC       Name=10;
CC         IsoId=O94833-1; Sequence=External;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Highly expressed in skeletal muscle and
CC       cultured keratinocytes.
CC   -!- DISEASE: BPAG1 is an autoantigen of bullous pemphigoid
CC       [MIM:600088], an autoimmune subepithelial skin blistering disease.
CC   -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC   -!- SIMILARITY: Contains 1 actin-binding domain.
CC   -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   -!- SIMILARITY: Contains 11 plectin repeats.
CC   -!- SIMILARITY: Contains 4 spectrin repeats.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M69225; -; NOT_ANNOTATED_CDS.
DR   EMBL; L11690; AAA52288.1; -.
DR   EMBL; U31850; AAC50243.1; -.
DR   EMBL; U31851; AAC50244.1; -.
DR   EMBL; AL096710; -; NOT_ANNOTATED_CDS.
DR   EMBL; AY032900; AAK63130.1; -.
DR   EMBL; AY032901; AAK63131.1; -.
DR   EMBL; M63618; AAA35606.1; -.
DR   EMBL; X58677; CAA41528.1; -.
DR   EMBL; M22942; AAA35538.1; -.
DR   EMBL; U04850; AAA57184.1; -.
DR   EMBL; U04850; AAA57185.1; -.
DR   PIR; I56317; A40937.
DR   Genew; HGNC:1090; BPAG1.
DR   MIM; 113810; -.
DR   MIM; 600088; -.
DR   GO; GO:0005604; C:basement membrane; TAS.
DR   GO; GO:0005737; C:cytoplasm; IEP.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; IEP.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; TAS.
DR   GO; GO:0045104; P:intermediate filament cytoskeleton organiza. . .; IEP.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR001101; Plectin_repeat.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00681; Plectin; 5.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00435; spectrin; 3.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00150; SPEC; 2.
DR   PROSITE; PS50002; SH3; FALSE_NEG.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR   PROSITE; PS50021; CH; 2.
KW   Antigen; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW   Structural protein; Cytoskeleton; Cell adhesion; Calcium;
KW   Calcium-binding; Alternative splicing.
FT   NON_TER       1      1
FT   DOMAIN        1   1669       GLOBULAR 1.
FT   DOMAIN     1670   2441       CENTRAL FIBROUS ROD DOMAIN.
FT   DOMAIN     2442   3214       GLOBULAR 2.
FT   DOMAIN      215    439       ACTIN-BINDING.
FT   DOMAIN      219    322       CH 1.
FT   DOMAIN      335    436       CH 2.
FT   REPEAT      774    851       SPECTRIN 1.
FT   REPEAT      876    959       SPECTRIN 2.
FT   REPEAT      984   1077       SPECTRIN 3.
FT   DOMAIN     1091   1143       SH3.

  Query Match             7.4%;  Score 95;  DB 1;  Length 3214;
  Best Local Similarity   25.4%;  Pred. No. 84;
  Matches   46;  Conservative   35;  Mismatches   62;  Indels   38;  Gaps    9;

Qy          2 AYKYDRDLEFLK--QLESSDLLDLF----EVLVFGKDGEKRHNEKLTSSIEYKRHGDDYA 55
              ::: :::|| |:  | :|  | : |    | |:     || :|:|:    |     :: |
Db       2075 SFRDEKELERLQICQRKSDHLKEQFEKSHEQLLQNIKAEKENNDKIQRLNEELEKSNECA 2134

Qy         56 KYAERIAEEL--QYYGSNSFASFIKGEG---VLYKEILCDVCDKLKV------------ 97
              :  ::  ||| |   :      |: |    || : :   |: ||:
Db       2135 EMLKQKVEELTRQNNETKLMMQRIQAESENIVLEKQTIQQRCEALKIQADGFKDQLRSTN 2194

Qy         98 -NYNKKTETTLIEQNMLSKI--LERSLEEMDD--EEVKEMCDELSI------KNTDNLNR 146
               : :|:|:|   ||:   ||  ||  | :  :   | |: ||: :|      |  |||
Db       2195 EHLHKQTKT---EQDFQRKIKCLEEDLAKSQNLVSEFKQKCDQQNIIIQNTKKEVRNLNA 2251

Qy        147 Q 147
              :
Db       2252 E 2252


RESULT 13
REST_CHICK
ID   REST_CHICK     STANDARD;      PRT;  1433 AA.
AC   O42184; O42228; O57563; O57564;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Restin (Cytoplasmic linker protein-170) (CLIP-170).
GN   RSN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98137792; PubMed=9469933;
RA   Griparic L., Volosky J.M., Keller T.C. III;
RT   "Cloning and expression of chicken CLIP-170 and restin isoforms.";
RL   Gene 206:195-208(1998).
RN   [2]
RP   SEQUENCE OF 17-1139 FROM N.A. (ISOFORMS 3 AND 4).
RC   TISSUE=Pectoralis muscle;
RX   MEDLINE=99002898; PubMed=9784600;
RA   Griparic L., Keller T.C. III;
RT   "Identification and expression of two novel CLIP-170/Restin isoforms
RT   expressed predominantly in muscle.";
RL   Biochim. Biophys. Acta 1405:35-46(1998).
CC   -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC       THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC       CYTOSKELETON (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=O42184-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O42184-2; Sequence=VSP_000761;
CC       Name=3; Synonyms=CLIP-170(11);
CC         IsoId=O42184-3; Sequence=VSP_000762, VSP_000763;
CC       Name=4; Synonyms=CLIP-170(11+35);
CC         IsoId=O42184-4; Sequence=VSP_000764;
CC   -!- SIMILARITY: Contains 2 CAP-Gly domains.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AF014012; AAC60344.1; -.
DR   EMBL; AF020764; AAC60345.1; -.
DR   EMBL; AF045650; AAC03547.1; -.
DR   EMBL; AF045651; AAC03548.1; -.
DR   InterPro; IPR000938; CAP-Gly.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01302; CAP_GLY; 2.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS00845; CAP_GLY_1; 2.
DR   PROSITE; PS50245; CAP_GLY_2; 2.
KW   Cytoskeleton; Microtubule; Coiled coil; Repeat; Alternative splicing.
FT   DOMAIN        79    121       CAP-GLY 1.
FT   DOMAIN       144    207       SER-RICH.
FT   DOMAIN       235    277       CAP-GLY 2.
FT   DOMAIN       305    332       SER-RICH.
FT   DOMAIN       351   1353       COILED COIL (POTENTIAL).
FT   DOMAIN      1414   1427       CCHC-BOX.
FT   VARSPLIC     458    492       Missing (in isoform 2).
FT                                 /FTId=VSP_000761.
FT   VARSPLIC     458    492       TQTKLEHARIKELEQSLLFEKTKADKLQRELEDTR -> RK
FT                                 RQISEDPEN (in isoform 3).
FT                                 /FTId=VSP_000762.
FT   VARSPLIC     803    803       S -> GGSSKVS (in isoform 3).
FT                                 /FTId=VSP_000763.
FT   VARSPLIC     458    458       T -> RKRQISEDPENT (in isoform 4).
FT                                 /FTId=VSP_000764.
FT   CONFLICT     309    309       K -> R (IN REF. 2; AAC03547).
FT   CONFLICT     440    440       E -> V (IN REF. 2; AAC03548).
SQ   SEQUENCE   1433 AA;  161026 MW;  5631CE8683498E23 CRC64;

  Query Match             7.4%;  Score 94.5;  DB 1;  Length 1433;
  Best Local Similarity   24.4%;  Pred. No. 34;
  Matches   52;  Conservative   36;  Mismatches   74;  Indels   51;  Gaps   10;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              |  |:| |::  ||    :|||  | |    | :|:::|  | : | |:  :  ||  :|
Db       1020 MITKHDADIKGFKQ----NLLDAEEAL---KAAQKKNDELETQAEELKKQAEQ-AKADKR 1071

Qy         61 IAEELQYYGSNSFASFIKGEGVLYK---EILCDVCDKLKVNYNKKTETTLIEQNMLSKIL 117
              | ||     :    | :  :::   | |  : :  : |   : |  :::|| |
Db       1072 AEEVLQ-----TMEKVTKEKDAIHQEKIETLASLENSRQTNEKLQNELDMLKQNNLKNEE 1126

Qy        118 E----RSLEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVI-VA 172
              |    : |  :::::|:|:  |           :|| |             || ||| :
Db       1127 ELTKSKELLNLENKKVEELKKEF----------EALKLAAAQ-------KSQQLAALQEE 1169

Qy        173 NAVAKTILGRG-------------LSLAGNQVL 192
              |     |||                 |:  ||:|
Db       1170 NVKLAEELGRSRDEVTSHQKLEEERSVLNNQLL 1202


RESULT 14
PAC_BACME
ID   PAC_BACME      STANDARD;      PRT;   802 AA.
AC   Q60136; Q9S463;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
DE   (Penicillin G amidohydrolase).
GN   PAC OR PGA.
OS   Bacillus megaterium.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1404;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 14945;
RA   Kang J.H., Kim S.J., Park Y.C., Hwang Y., Yoo O.J., Kim Y.C.;
RT   "Nucleotide sequence of the penicillin G acylase gene from
RT   Bacillus megaterium and characteristics of the enzyme.";
RL   Misainmurhag Hoiji 32:215-221(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 14945;
RX   MEDLINE=95180705; PubMed=7875576;
RA   Martin L.M., Prieto A.M., Cortes E., Garcia J.L.;
RT   "Cloning and sequencing of the pac gene encoding the penicillin G
RT   acylase of Bacillus megaterium ATCC 14945.";
RL   FEMS Microbiol. Lett. 125:287-292(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CA4098;
RX   PubMed=12114980;
RA   Yang S., Huang Y.H., Huang X.D., Li S.Y., Yuan Z.Y.;
RT   "High expression of penicillin G acylase gene from Bacillus megaterium
RT   in Bacillus subtilis.";
RL   Acta Biochim. Biophys. Sin. 31:601-603(1999).
CC   -!- CATALYTIC ACTIVITY: Penicillin + H(2)O = a fatty acid anion + 6-
CC       aminopenicillanate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (Potential).
CC   -!- SUBUNIT: Heterodimer of an alpha chain and a beta chain (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Extracellular (Potential).
CC   -!- SIMILARITY: Belongs to peptidase family S45.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; U07682; AAB41343.1; -.
DR   EMBL; Z37542; CAA85774.1; -.
DR   EMBL; AF161313; AAD45609.1; -.
DR   PIR; S49252; S49252.
DR   HSSP; P06875; 1AJQ.
DR   MEROPS; S45.001; -.
DR   InterPro; IPR002692; Peptidase_S45.
DR   Pfam; PF01804; Penicil_amidase; 1.
KW   Hydrolase; Antibiotic resistance; Zymogen; Calcium-binding; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    802       PENICILLIN G ACYLASE ZYMOGEN.
FT   CHAIN        25    234       PENICILLIN G ACYLASE ALPHA SUBUNIT.
FT   PROPEP      235    265       SPACER PEPTIDE.
FT   CHAIN       266    802       PENICILLIN G ACYLASE BETA SUBUNIT.
FT   ACT_SITE    266    266       BY SIMILARITY.
FT   METAL       177    177       CALCIUM (POTENTIAL).
FT   METAL       341    341       CALCIUM (POTENTIAL).
FT   VARIANT       3      3       T -> M (IN STRAIN CA4098).
FT   VARIANT     224    224       D -> E (IN STRAIN CA4098).
FT   VARIANT     232    232       I -> K (IN STRAIN CA4098).
FT   VARIANT     254    254       T -> S (IN STRAIN CA4098).
FT   VARIANT     349    349       A -> T (IN STRAIN CA4098).
FT   VARIANT     470    470       S -> N (IN STRAIN CA4098).
FT   VARIANT     524    524       F -> Y (IN STRAIN CA4098).
FT   VARIANT     569    569       Q -> P (IN STRAIN CA4098).
FT   VARIANT     586    586       I -> A (IN STRAIN CA4098).
FT   VARIANT     657    657       N -> S (IN STRAIN CA4098).
FT   VARIANT     740    740       T -> K (IN STRAIN CA4098).
FT   VARIANT     789    791       NKA -> YKS (IN STRAIN CA4098).
SQ   SEQUENCE   802 AA;  91987 MW;  877CA0564E50DFBD CRC64;

  Query Match             7.3%;  Score 93.5;  DB 1;  Length 802;
  Best Local Similarity   21.8%;  Pred. No. 20;
  Matches   62;  Conservative   39;  Mismatches  107;  Indels   77;  Gaps   10;

Qy          4 KYDRDLEFLKQLESSDLLDLFEV--LVFGKDG------EKRHNEKLTSSIEYKRHGDDYA 55
              |  ||    :  | : :  ||:|    |  ||          :| ||   | |      |:||
Db         33 KVVRDNFGVPHLYAKNKKDLYEAYGYVMAKDRLFQLEMFRRGNEGTVSEI----FGEDYL 88

Qy         56 ------------KYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYN-- 100
```

```
                     |  ::: : |       | | :|      | | |  |||   ::
Db         89 SKDEQSRRDGYSNKEIKKMIDGLDRQPKELIAKFAEGISRYVNEALKDPDDKLSKEFHEY 148

Qy        101 -----KKTETTLIEQNMLS-------------KILERSLEEMDDEEVKEMCDELSIKN- 140
                 | | | ::   |:|               :|| :   |   |  ::| |:|  ||
Db        149 QFLPQKWTSTDVVRVYMVSMTYFMDNHQELKNAEILAKLEHEYGTEVSRKMFDDLVWKND 208

Qy        141 ---------------TDNLNRQALSAATLTLFKMGGFK-------SYQLAVIVANAVAK 177
                            : : : | ||:| :   :  | :       | :| : :
Db        209 PSAPTSIVSEGKPKRDSSSQSLQILSSAVIKASEKVGKERENFVQTSEELGLPLKIGSNA 268

Qy        178 TILGRGLSLAGNQVLTRTLSFLTGP-VGWIITGVWTAIDIAGPAY 221
               |:|   |  || :|       :|| ||::  |    : :  | :
Db        269 AIVGSEKSATGNALL------FSGPQVGFVAPGFLYEVGLHAPGF 307
```

```
RESULT 15
PAC_ARTVI
ID   PAC_ARTVI      STANDARD;      PRT;   802 AA.
AC   P31956;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
DE   (Penicillin G amidohydrolase).
GN   PAC OR PA.
OS   Arthrobacter viscosus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Micrococcineae; Micrococcaceae; Arthrobacter.
OX   NCBI_TaxID=1673;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15294;
RX   MEDLINE=94259306; PubMed=8200542;
RA   Konstantinovic M., Marjanovic N., Ljubijankic G., Glisin V.;
RT   "The penicillin amidase of Arthrobacter viscosus (ATCC 15294).";
RL   Gene 143:79-83(1994).
CC   -!- CATALYTIC ACTIVITY: Penicillin + H(2)O = a fatty acid anion + 6-
CC       aminopenicillanate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (Potential).
CC   -!- SUBUNIT: Heterodimer of an alpha chain and a beta chain (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Extracellular (Potential).
CC   -!- SIMILARITY: Belongs to peptidase family S45.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L04471; AAA22077.1; -.
DR   PIR; I39665; I39665.
DR   HSSP; P05875; 1AJQ.
DR   MEROPS; S45.001; -.
DR   InterPro; IPR002692; Peptidase_S45.
DR   Pfam; PF01804; Penicil_amidase; 1.
KW   Hydrolase; Antibiotic resistance; Zymogen; Calcium-binding; Signal.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25    802       PENICILLIN G ACYLASE ZYMOGEN.
FT   CHAIN       25    234       PENICILLIN G ACYLASE ALPHA SUBUNIT.
FT   PROPEP     235    265       SPACER PEPTIDE.
FT   CHAIN      266    802       PENICILLIN G ACYLASE BETA SUBUNIT.
FT   ACT_SITE   266    266       BY SIMILARITY.
FT   METAL      177    177       CALCIUM (POTENTIAL).
FT   METAL      341    341       CALCIUM (POTENTIAL).
SQ   SEQUENCE   802 AA;  92113 MW;  9863E58C526C85D7 CRC64;

  Query Match             7.2%;  Score 92.5;  DB 1;  Length 802;
  Best Local Similarity   21.8%;  Pred. No. 24;
  Matches   62;  Conservative   39;  Mismatches  107;  Indels   77;  Gaps   10;
```

```
Qy          4 KYDRDLEFLKQLESSDLLDLFEV--LVFGKDG------EKRHNEKLTSSIEYKRHGDDYA 55
                |  ||    :  | : :  ||:|    |  ||        :| ||   | |     |:||
Db         33 KVVRDNFGVPHLYAKNKKDLYEAYGYVMAKDRLFQLEMFRRGNEGTVSEI----FGEDYL 88

Qy         56 -------------KYAERIABELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYN-- 100
                         |  ::: : |       | | :|      | | |  |||   ::
Db         89 SKDEQSRRDGYSNKEIKKMIDGLDRQPRELIAKFAEGISRYVNEALKDPDDKLSKEFHEY 148

Qy        101 -----KKTETTLIEQNMLS-------------KILERSLEEMDDEEVKEMCDELSIKN- 140
                 | | | ::   |:|               :|| :   |   |  ::| |:|  ||
Db        149 QFLPQKWTSTDVVRVYMVSMTYLWIITRELKNAEILAKLEHEYGTEVSRKMFDDLVWKND 208

Qy        141 ---------------TDNLNRQALSAATLTLFKMGGFK-------SYQLAVIVANAVAK 177
                            : : : | ||:| :   :  | :       | :| : :
Db        209 PSAPTSIVSEGKPKRESSSQSLQKLSSAVIKASEKVGKERENFVQSSEELGLPLKIGSNA 268

Qy        178 TILGRGLSLAGNQVLTRTLSFLTGP-VGWIITGVWTAIDIAGPAY 221
               |:|   |  || :|       :|| ||::  |    : :  | :
Db        269 AIVGSEKSATGNALL------FSGPQVGFVAPGFLYEVGLHAPGF 307
```

Search completed: July 5, 2004, 03:12:51
Job time : 12 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 5, 2004, 03:06:48 ; Search time 29 Seconds
(without alignments)
2752.624 Million cell updates/sec

Title: US-09-732-091-4
Perfect score: 1279
Sequence: 1 MAYKYDRDLEFLKQLESSDL.....LRLKTQQANGDKKSLQIESI 253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	265	20.7	237	16	Q8XGV3	Q8xgv3 salmonella
2	260	20.3	237	16	Q83SR4	Q83sr4 shigella fl
3	259	20.3	237	16	Q8FLC7	Q8flc7 escherichia
4	133	10.4	39	16	O26108	O26108 helicobacte
5	105	8.2	1245	5	Q8IAN1	Q8ian1 plasmodium
6	103.5	8.1	403	16	Q8R9Z4	Q8r9z4 thermoanaer
7	102.5	8.0	276	16	Q8DIH4	Q8dih4 synechococc
8	102	8.0	1590	5	Q8MX30	Q8mx30 naegleria g
9	101	7.9	1455	5	Q8IKG8	Q8ikg8 plasmodium
10	100.5	7.9	726	13	P87397	P87397 oncorhynchu
11	99.5	7.8	556	16	Q7VHM0	Q7vhm0 helicobacte
12	99.5	7.8	633	16	Q882L0	Q882l0 pseudomonas
13	99	7.7	202	5	Q8IJZ2	Q8ijz2 plasmodium
14	99	7.7	222	4	O43633	O43633 homo sapien
15	99	7.7	222	11	Q9DB34	Q9db34 mus musculu
16	99	7.7	723	13	Q7ZZU9	Q7zzu9 astyanax fa
17	98.5	7.7	1009	16	Q99UD0	Q99ud0 staphylococ
18	98	7.7	282	16	Q8ZCL9	Q8zcl9 yersinia pe
19	98	7.7	413	9	Q8SDW1	Q8sdw1 bacteriopha
20	98	7.7	1099	16	Q98QT2	Q98qt2 mycoplasma
21	98	7.7	2359	5	Q8I297	Q8i297 plasmodium
22	98	7.7	2849	5	Q8IHY4	Q8ihy4 plasmodium
23	97.5	7.6	311	16	Q9PIC9	Q9pic9 campylobact
24	97.5	7.6	728	16	Q9PI94	Q9pi94 campylobact
25	97.5	7.6	792	10	Q9ARQ9	Q9arq9 oryza sativ
26	96.5	7.5	1102	16	Q827Q7	Q827q7 streptomyce
27	96.5	7.5	1341	5	Q8IEQ5	Q8ieq5 plasmodium
28	96	7.5	262	5	O61768	O61768 caenorhabdi
29	96	7.5	420	16	Q97MM5	Q97mm5 clostridium
30	96	7.5	4405	5	Q8ILZ2	Q8ilz2 plasmodium
31	95.5	7.5	944	17	Q8TPC2	Q8tpc2 methanosarc
32	94.5	7.4	406	16	Q8RH57	Q8rh57 fusobacteri
33	94	7.3	385	2	Q9ZI88	Q9zi88 borrelia bu
34	94	7.3	576	10	Q84LB2	Q84lb2 malus domes
35	93.5	7.3	1413	5	Q8I589	Q8i589 plasmodium
36	93	7.3	355	17	Q97UF7	Q97uf7 sulfolobus
37	93	7.3	610	16	O51577	O51577 borrelia bu
38	93	7.3	1173	2	Q9KHU3	Q9khu3 helicobacte
39	92.5	7.2	391	9	Q7Y5E5	Q7y5e5 aeromonas p
40	92.5	7.2	430	5	O61767	O61767 caenorhabdi
41	92.5	7.2	1960	10	Q8S6N2	Q8s6n2 oryza sativ
42	92	7.2	308	16	Q9PPG7	Q9ppg7 campylobact
43	92	7.2	309	16	Q8E6J7	Q8e6j7 streptococc
44	92	7.2	309	16	Q8E143	Q8e143 streptococc
45	92	7.2	331	17	Q96XT6	Q96xt6 sulfolobus

ALIGNMENTS

RESULT 1
Q8XGV3
ID Q8XGV3 PRELIMINARY; PRT; 237 AA.
AC Q8XGV3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Positive regulator for sigma H (Sigma 32)
DE promoters, permitting growth at high temperature).
GN T0010 OR HTGA OR STM0010 OR STY0010.
OS Salmonella typhi, and
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601, 602;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;

```
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AE016834; AAO67744.1; -.
DR   EMBL; AE008693; AAL18974.1; -.
DR   EMBL; AL627265; CAD01163.1; -.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   237 AA;  26515 MW;  184ADE026EA5BCA9 CRC64;

  Query Match             20.7%;  Score 265;  DB 16;  Length 237;
  Best Local Similarity   32.6%;  Pred. No. 2.6e-12;
  Matches   79;  Conservative   40;  Mismatches  101;  Indels   22;  Gaps    7;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKR------HNEKLTSSIEYKRHGDDY 54
              : | :| | :||:       | |    :|   : |: |       ||| |  ::|   | : :
Db          3 VTYLHDEDLDFLQHCSEEQLADFARLLTHNEKGKARLSSVLSHNE-LFKAME--GHPEQH 59

Qy         55 AKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLS 114
               :  : || | |:|| :| |: ::| |  |: || ||  :||:  :|   | ||| :|
Db         60 RRNWQLIAGEFQHYGGDSIANKLRGHGKQYRAILLDVAKRLKLKADKSMSTFEIEQQLLE 119

Qy        115 KILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFK-MGGFKSYQLAVIVAN 173
                |  : ::||    :|   :  | ::      |     :| | :    | ||  |:
Db        120 HFLRHTWQKMDAAHKQEFLQAVDAKVSELEELLPLLMKDRSLAKGVSHLLSTQLTRILRT 179

Qy        174 AVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVAT 233
              | :||| || | |         : | ||||  : ||     ::| |||||||| : :|
Db        180 HAAMSILGHGL-LRG--------AGLGGPVGAALNGVKA---MSGSAYRVTIPAVLQIAC 227

Qy        234 LR 235
               ||
Db        228 LR 229


RESULT 2
Q83SR4
ID   Q83SR4      PRELIMINARY;      PRT;   237 AA.
AC   Q83SR4;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative oxidoreductase.
GN   SF0012 OR S0012.
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Shigella.
OX   NCBI_TaxID=623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=301 / Serotype 2a;
RX   MEDLINE=22272406; PubMed=12384590;
RA   Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA   Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA   Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA   Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA   Yu J.;
RT   "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT   through comparison with genomes of Escherichia coli K12 and O157.";
RL   Nucleic Acids Res. 30:4432-4441(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX   MEDLINE=22590274; PubMed=12704152;
RA   Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA   Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA   Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA   Schwartz D.C., Blattner F.R.;
RT   "Complete genome sequence and comparative genomics of Shigella
RT   flexneri serotype 2a strain 2457T.";
RL   Infect. Immun. 71:2775-2786(2003).
DR   EMBL; AE015039; AAN41678.1; -.
DR   EMBL; AE016978; AAP15557.1; -.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Complete proteome.
SQ   SEQUENCE   237 AA;  26709 MW;  95509A1BCB8B4CF5 CRC64;

  Query Match             20.3%;  Score 260;  DB 16;  Length 237;
  Best Local Similarity   30.8%;  Pred. No. 6.3e-12;
  Matches   76;  Conservative   44;  Mismatches   91;  Indels   36;  Gaps    6;

Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKR---HGDDYAKYAE 59
              |  | ||:||:       | :   :|   : |: | :  |  : :|    | : : :  :
Db          5 YLNDSDLDFLQHCSEEQLANFARLLTHNEKGKTRLSSILMRNELFKSMEGHPEQHRRNWQ 64

Qy         60 RIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119
               || |||::| :| | : ::| | ||: || ||  :||:  :|:  |  ||| :| : |
Db         65 LIAGELQHFGGDSIANKLRGHGKLYRAILLDVSKRLKLKADKEMSTFEIEQQLLEQFLRN 124

Qy        120 SLEEMDDE----------EVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLA 168
              : ::||:|           | |: : | :   | |  : :|              | ||
Db        125 TWKKMDEEHKQEFLHAVDARVNELEELLPLLMKDKLLAKGVS----------HLLSSQLT 174

Qy        169 VIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPAC 228
              |:    | ::|| || | |         : | ||||  : ||     ::|  |||||||
Db        175 RILRTHAAMSVLGHGL-LRG--------AGLGGPVGAALNGVKA---VSGSTYRVTIPAV 222

Qy        229 IVVATLR 235
              : :| ||
Db        223 LQIACLR 229


RESULT 3
Q8FLC7
ID   Q8FLC7      PRELIMINARY;      PRT;   237 AA.
AC   Q8FLC7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein yaaW.
GN   YAAW OR C0016.
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR   EMBL; AE016755; AAN78516.1; -.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   237 AA;  26727 MW;  B8C190712375B31D CRC64;

  Query Match             20.3%;  Score 259;  DB 16;  Length 237;
  Best Local Similarity   30.8%;  Pred. No. 7.5e-12;
  Matches   76;  Conservative   44;  Mismatches   91;  Indels   36;  Gaps    6;
```

```
Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKR---HGDDYAKYAE 59
              |  | |||:||:       | :   :|   : |: | :  |  :  :|     | : : :  :
Db          5 YLNDSDLDFLQHCSEEQLANFARLLTHNEKGKTRLSSVLMRNELFKSMEGHPERHRRNWQ 64

Qy         60 RIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119
              || | |::| :| |: ::| | ||: || ||  :||:  :|:  |  ||| :| : |
Db         65 LIAGEFQHFGGDSIANKLRGHGKLYRAILLDVSKRLKLKADKEMSTFEIEQQLLEQFLRN 124

Qy        120 SLEEMDDE-----------EVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLA 168
              : ::|| :|           | |: : | :   | |  : :|              | ||
Db        125 TWKKMDEEHKQEFLHAVDARVNELEELLPLLMKDKLLAKGVS----------HLLSSQLT 174

Qy        169 VIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPWGWIITGVWTAIDIAGPAYRVTIPAC 228
              |:    | ::|| || | |         : | |||  : ||    ::| ||||||||
Db        175 RILRTHAAMSVLGHGL-LRG--------AGLGGPVGAALNGVKA---VSGSAYRVTIPAV 222

Qy        229 IVVATLR 235
              : :| ||
Db        223 LQIACLR 229
```

RESULT 4
O26108
ID O26108 PRELIMINARY; PRT; 39 AA.
AC O26108;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein HP1590.
GN HP1590.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
DR EMBL; AE000656; AAD08629.1; -.
DR PIR; F64718; F64718.
DR TIGR; HP1590; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 39 AA; 4636 MW; EE53CF01F4C1C909 CRC64;

Query Match 10.4%; Score 133; DB 16; Length 39;
Best Local Similarity 66.7%; Pred. No. 0.0028;
Matches 26; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNE 39
              |||:|| |||||:| |||| |||: ||: :||  | ||
Db          1 MAYRYDSDLEFLKRLSSSDLKDLFDALVYDEDGTLRMNE 39
```

RESULT 5
Q8IAN1
ID Q8IAN1 PRELIMINARY; PRT; 1245 AA.
AC Q8IAN1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF08_0127.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL844507; CAD51332.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1245 AA; 147911 MW; D856486AFDFE4DDF CRC64;

Query Match 8.2%; Score 105; DB 5; Length 1245;
Best Local Similarity 26.5%; Pred. No. 32;
Matches 35; Conservative 21; Mismatches 48; Indels 28; Gaps 5;

```
Qy         34 EKRHNEKLTSSIEYKRHGD---DYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCD 90
              ||:  ||:   : |  | |   |   | :  ||:     :      ||     ||
Db        542 EKKEEEKIMKKL-YSEHPDLVKDLMNYEKEYAEKRNLLNNIKKRKTIKTLN-RYKN---- 595

Qy         91 VCDKLKVNYNKKTETTLIEQNMLSKILERSL----------------EEMDDEEVKEMC 133
                ::||    |  ::   |:||| ||:|::                :: ||: : :
Db        596 --EQLKKEILKSIQSNKEEKNMLKKIIEKTALTEQTNEDNNMMIHDDNDDDDDDNMSEDT 653

Qy        134 DELSIKNTDNLN 145
              ||:| |: ||:|
Db        654 DEISSKDHDNMN 665
```

RESULT 6
Q8R9Z4
ID Q8R9Z4 PRELIMINARY; PRT; 403 AA.
AC Q8R9Z4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein TTE1436.
GN TTE1436.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013102; AAM24658.1; -.
DR GO; GO:0009424; C:flagellar hook (sensu Bacteria); IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0009296; P:flagella biogenesis; IEA.
DR InterPro; IPR001635; Flag_hook.
DR Pfam; PF02120; Flg_hook; 1.
DR PRINTS; PR01007; FLGHOOKFLIK.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 403 AA; 46116 MW; 60C22B9FBB51A56F CRC64;

Query Match 8.1%; Score 103.5; DB 16; Length 403;
Best Local Similarity 20.1%; Pred. No. 9.8;
Matches 63; Conservative 47; Mismatches 98; Indels 105; Gaps 12;

```
Qy         10 EFLKQLESSDLLDLFEVLVFGKDGEKRHNEK----------------LTSSIE---YKRH 50
              ||| |      :::|:        ||   :|                :| |:|    ::
Db         61 EFLSQKNGLKNVEIFQ--------EKVKEDKDIMEDLNNFIPALMQTITLSLETISSEKG 112

Qy         51 GDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQ 110
               |: |  |::  ||        |||     :|||   : | || |:| :    :||:
```

```
Db        113 ASDFEKVREKLEVALQ--------GFIKERNFTFKEIAKKISDFLKENFNIELSPEVIER 164

Qy        111 NM-LSKILERSLEEMDDEEVKEMCDELSIKNTD-----NLNRQALSAA------------ 152
              :: |:|: :     : |   |:  :|   || |      ::::|  ||
Db        165 HIKLAKVKDLDKPFLQDLNQKDFAEENVQKNQDKTSQLKIDKEAFIAAKEAKEEKTEKKS 224

Qy        153 -----TLTLFKMGGFKSYQLAVIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWII 207
                  ||  |     :: :  |:: |:           |  : |                |:
Db        225 FDVKQEFVFFKNEG---KPVSNLTYNSIKKS----------NDPVDRLFR-------QIV 264

Qy        208 TGVWTAIDIAGPAYRVTIPACI--------------VVATL-------------RLKTQQ 240
                |: | :    :  | :   |                :|||:             |   |
Db        265 DNVFVAKEKGASSVTVNLKPEILGKLQISLKSIDGNIVATIVTESEKTKHQIESNLSLLQ 324

Qy        241 ANGDKKSLQIESI 253
              |  | | ::|||:
Db        325 AQLDLKGIKIESV 337


RESULT 7
Q8DIH4
ID   Q8DIH4      PRELIMINARY;      PRT;   276 AA.
AC   Q8DIH4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Tll1614 protein.
GN   TLL1614.
OS   Synechococcus elongatus (Thermosynechococcus elongatus).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=32046;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP-1;
RX   MEDLINE=22225144; PubMed=12240834;
RA   Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA   Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the thermophilic cyanobacterium
RT   Thermosynechococcus elongatus BP-1.";
RL   DNA Res. 9:123-130(2002).
DR   EMBL; AP005374; BAC09166.1; -.
KW   Complete proteome.
SQ   SEQUENCE   276 AA;  31282 MW;  D0577A9D97E0CA92 CRC64;

  Query Match             8.0%;  Score 102.5;  DB 16;  Length 276;
  Best Local Similarity   19.0%;  Pred. No. 7.2;
  Matches   50;  Conservative   53;  Mismatches  111;  Indels   49;  Gaps   10;

Qy         12 LKQLESSDLLDLFEVLVFGKDGEKRHN--EKLTSSIEYKRHGDDYAKYAERIAEELQYYG 69
              |:     :| || |:|       :| |  : ||:          |    : : | |  ::
Db         24 LELATEEELQDLTEILF-----RRRLNPLDYLTTPDPIAVQAQDRQAWLDDIEERFRFLA 78

Qy         70 SNSFASFIKGEG--VLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSLEEMDDE 127
              ::   : :||:   : |:: |  ||  ||: ::       :|  :   :|:|  ::: |:
Db         79 ADGL-TVLKGKAQQISYRQTLMRVCRYLKIKFSPSWTVPELEMEIFLNVLQRMWKKLGDQ 137

Qy        128 EVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL------- 180
              : : :  ::     :  :   :|   : |   ||       | |  ::| ::::
Db        138 DRRVLAAQIQESLPELHHGHPISMEMVRLVLEGG------AAIAISSVVRSMVVQQVARQ 191

Qy        181 ------GRGLSLA-----------GNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPA--- 220
                   | ||:|                : |::       |     ||    |:   |
Db        192 FAIRFAGSKLSIAPLVSRGAAMGVARLAVGRSILAFVSTALW----VWFIADLGWQAIST 247

Qy        221 -YRVTIPACIVVATLR-LKTQQA 241
               |    ||     :| :|  |: :||
Db        248 NYARIIPTIFAIAQIRLLRGEQA 270


RESULT 8
Q8MX30
ID   Q8MX30      PRELIMINARY;      PRT;  1590 AA.
AC   Q8MX30;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RNA polymerase II largest subunit (Fragment).
GN   RPB1.
OS   Naegleria gruberi.
OC   Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.
OX   NCBI_TaxID=5762;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12032239;
RA   Dacks J.B., Marinets A., Ford Doolittle W., Cavalier-Smith T.,
RA   Logsdon J.M. Jr.;
RT   "Analyses of RNA Polymerase II Genes from Free-Living Protists:
RT   Phylogeny, Long Branch Attraction, and the Eukaryotic Big Bang.";
RL   Mol. Biol. Evol. 19:830-840(2002).
DR   EMBL; AF395110; AAM45151.1; -.
DR   GO; GO:0005665; C:DNA-directed RNA polymerase II, core complex; IEA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0006366; P:transcription from Pol II promoter; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR006592; RNA_polA_N.
DR   InterPro; IPR000684; RNA_polII_repeat.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR007080; RNA_pol_Rpb1_1.
DR   InterPro; IPR007066; RNA_pol_Rpb1_3.
DR   InterPro; IPR007083; RNA_pol_Rpb1_4.
DR   InterPro; IPR007081; RNA_pol_Rpb1_5.
DR   InterPro; IPR007075; RNA_pol_Rpb1_6.
DR   InterPro; IPR007073; RNA_pol_Rpb1_7.
DR   Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR   Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR   Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR   Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR   Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR   Pfam; PF04992; RNA_pol_Rpb1_6; 1.
DR   Pfam; PF04990; RNA_pol_Rpb1_7; 1.
DR   Pfam; PF05001; RNA_pol_Rpb1_R; 5.
DR   SMART; SM00663; RPOLA_N; 1.
DR   PROSITE; PS00115; RNA_POL_II_REPEAT; 7.
FT   NON_TER       1      1
FT   NON_TER    1590   1590
SQ   SEQUENCE   1590 AA;  177446 MW;  5F4C7959D4D77D6B CRC64;

  Query Match             8.0%;  Score 102;  DB 5;  Length 1590;
  Best Local Similarity   22.8%;  Pred. No. 73;
  Matches   61;  Conservative   37;  Mismatches   71;  Indels   98;  Gaps   14;

Qy         28 VFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEELQYYGS-NSFASF---------I 77
              ::|:||    |  : : ||:      |  ||: ::      :| |  |  ||         :
Db        810 LYGEDGMDAHKVE-SQSIDIMNLSD--AKFQDK-----YWYNSLGSSPSFTNPTESSVSM 861

Qy         78 KGEGVL---------YKEILCDVCDKL-KVNYNKKT----------------ETTLIEQN 111
                |:|          | ||| |    | ||    ||                    |  ::
Db        862 PSGGILKESPISTKVYDEILKDPATYLSKVREEYKTLMEDRHILRTEIFPNAENKIVMPV 921

Qy        112 MLSKILERSLEEMD----------------DEEVKEMCDEL-SIKNTDNLNRQALSAATL 154
              | :|:: : :|                  |::|::||:|  || || |: :| : |||
Db        922 NLKRIIKNAQKEYGIHPSIGKPSDMNPIYVIEKIKKLCDDLVVIKGTDELSTEAQTNATL 981

Qy        155 TLFKMGGFKSYQLAVIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWII------- 207
              || |              :  |   :| ||    :| : |:|       |::
Db        982 -LFSM-------------FLRSTFFFKGCSLKNCVLLKKPLTF------WLVKLELRFH 1020

Qy        208 ---------TGVWTAIDIAGPAYRVTI 225
                        |   |  |  || ::|:
Db       1021 NCLAQPGEMVGSVAAQSIGEPATQMTL 1047
```

```
RESULT 9
Q8IKG8
ID   Q8IKG8      PRELIMINARY;      PRT;  1455 AA.
AC   Q8IKG8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Rhoptry protein, putative.
GN   PF14_0637.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014826; AAN37250.1; -.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.
DR   InterPro; IPR002114; HPr_SerP_S.
DR   PROSITE; PS00589; PTS_HPR_SER; 1.
SQ   SEQUENCE   1455 AA;  174305 MW;  36FAA1760A898F99 CRC64;

  Query Match             7.9%;  Score 101;  DB 5;  Length 1455;
  Best Local Similarity   25.5%;  Pred. No. 78;
  Matches   36;  Conservative   29;  Mismatches   52;  Indels   24;  Gaps    6;

Qy          6 DRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSI-EYKRHGDDYAKYAERIAEE 64
              :|:| :|  :|     :|    :  :: |  : ||| : | ||::  ::     |::
Db        333 ERELYYLGMIE-----ELRNEIKTKEENEGNNIEKLENKIHEYEKQNEELRNEKEKLQST 387

Qy         65 LQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYN---KKTETTLIEQNMLSKILERSL 121
              :  | |::| :      :  ||     |::|| |||   :| |    || :  |
Db        388 INEY-SHNFNNLNDHNKITNKE-----CEELKNNYNTIKEKYERLKEEQEIYIK------ 435

Qy        122 EEMDDEEVKEMCDELSIKNTD 142
                :|| | : |||  :| :
Db        436 ---QEEEYKSLLDELENENNE 453


RESULT 10
P87397
ID   P87397      PRELIMINARY;      PRT;  726 AA.
AC   P87397;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Heat shock protein hsp90.
OS   Oncorhynchus tschawytscha (Chinook salmon) (King salmon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=74940;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99262176; PubMed=10329464;
RA   Palmisano A.N., Winton J.R., Dickhoff W.W.;
RT   "Sequence features and phylogenetic analysis of the stress protein
RT   hsp90alpha in chinook salmon (Oncorhynchus tshawytscha), a
RT   poikilothermic vertebrate.";
RL   Biochem. Biophys. Res. Commun. 258:784-791(1999).
DR   EMBL; U89945; AAB49983.1; -.
DR   HSSP; P07900; 1BYQ.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003754; F:chaperone activity; IEA.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR001404; Hsp90.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00183; HSP90; 1.
DR   PRINTS; PR00775; HEATSHOCK90.
DR   SMART; SM00387; HATPase_c; 1.
DR   PROSITE; PS00298; HSP90; 1.
SQ   SEQUENCE   726 AA;  83505 MW;  1A2A962C471CBCF2 CRC64;

  Query Match             7.9%;  Score 100.5;  DB 13;  Length 726;
  Best Local Similarity   21.6%;  Pred. No. 35;
  Matches   61;  Conservative   39;  Mismatches   81;  Indels  101;  Gaps   11;

Qy         21 LDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEELQYYGSNSFASFIKGE 80
              :|||  |   ||  |:  |: : :|:   | |  |:  ::::: |:|| ||| |     |
Db        414 MDLFVELSEDKDNYKKFYEQFSKNIKLGIHED--AQNRKKLSDMLRYYTSNSNAD----E 467

Qy         81 GVLYKEILCDVCDKLKVNYNKKTETT--------------------LIE---------- 109
              |  || :  : |  |  |    ||                      :||
Db        468 MVSLKEYVSRMKDTQKHIYYITGETKEQVANSSFVERLRKAGLEVIYMIEPIDEYCVQQL 527

Qy        110 -----QNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKS 164
                  :|::| : :  ||  :||: |:  :||: |  :|| :
Db        528 KEYDGKNLVS-VTKEGLELPEDEDEKKKQEELNTK-FENLCK----------------- 567

Qy        165 YQLAVIVANAVAKTILGRGLS--LAGNQVLTRTLSFLTGPVGWIIT------------- 208
                      | || : :     |:::   :|   ||
Db        568 ----------TMKDILDKKIEKVSVSNRLVSSPCCIVTSTYGWTANMERIMKSQALRDNS 617

Qy        209 --GVWTA---IDIAGPAYRVTIPACIVVATLRLKTQQANGDK 245
                |  ||   ::|         |   :| ||| | :    ||
Db        618 TMGYMTAKKHLEIN--------PTHPIVETLREKAEADKNDK 651


RESULT 11
Q7VHM0
ID   Q7VHM0      PRELIMINARY;      PRT;  556 AA.
AC   Q7VHM0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   HH0946.
OS   Helicobacter hepaticus.
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=32025;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 51449 / 3B1;
RX   MEDLINE=22709201; PubMed=12810954;
RA   Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA   Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA   Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA   Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT   "The complete genome sequence of the carcinogenic bacterium
RT   Helicobacter hepaticus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR   EMBL; AE017147; AAP77543.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   556 AA;  61984 MW;  0E48CF5376BAC48D CRC64;

  Query Match             7.8%;  Score 99.5;  DB 16;  Length 556;
  Best Local Similarity   23.8%;  Pred. No. 30;
  Matches   53;  Conservative   33;  Mismatches  104;  Indels   33;  Gaps    6;

Qy         32 DGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDV 91
              |  ||| | |   ||  ::  :                 |:     | |  | |    ::
Db        199 DDYKRHKENLEKMIEKGKNNPE--------------NQNAVEKAQKAEKALEKLNANNM 243
```

```
Qy          92 CDKLKVNYNKKTETTL--------IEQNMLSKILERSLEEMDDE---EVKEMCD-----E 135
Db         244 CNRLMCENPRITAVATQSIVASGHIAQAGFSDAIIAALATLANGVIWEVKDMFEGSIDTE 303

Qy         136 LSI-KNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGLSLAGNQVLTR 194
Db         304 TSILKRIKRLLKRTIEAFQATFGRGAGFGAIDAAVGVVGQIFRSIAGK-LKLVWDKIRTA 362

Qy         195 TLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLK 237
Db         363 LKSIYNGIVSYIKGEVSNLRELLGIILKSLFSAAWVVSTLALE 405
```

RESULT 12
Q882L0
ID Q882L0 PRELIMINARY; PRT; 633 AA.
AC Q882L0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN PSPTO2616.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE016865; AAO56120.1; -.
DR TIGR; PSPTO2616; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 1.
KW Complete proteome.
SQ SEQUENCE 633 AA; 68090 MW; D15A8BC00AC58637 CRC64;

Query Match 7.8%; Score 99.5; DB 16; Length 633;
Best Local Similarity 23.7%; Pred. No. 35;
Matches 69; Conservative 50; Mismatches 87; Indels 85; Gaps 18;

```
Qy           3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSS-----IEYKRHGDD---- 53
Db          70 YDAERASDWLKALERLEAHVKYAQEVF----DSPLNIPLVNAAAEALVEYRVHYDNLMRA 125

Qy          54 ----------YAKYAERIAEELQYYGSNSFASFIKGEG----------VLYKEILCDVCD 93
Db         126 TAAREATRGAFGQYADAGAEDLQKL--NAFARSDEGSASQRDAIVQAMTLFQKMRFDL-- 181

Qy          94 KLKVNYNKKTETTLIEQNMLSKILE--RSLEEMDDEE--VKEMCDEL-SIKNTDN---LN 145
Db         182 -RGYTYSLKAENRAPAEASMSAVINFVKGLQGFDSQSATIKHLVDSMVSYQNTMNQFTAA 240

Qy         146 RQALSAATLTLFKMGG--FKSY-QLAVIVANAV---------AKTILGRGLSLAGNQVLT 193
Db         241 QASIDQAQAGITKVIGVLFKSADQLS---ANQVSLRIEDVDQARTLLSVWL------VAA 291

Qy         194 RTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQANGD 244
Db         292 LIMSFL---AAWVITRL-----IVGP----------LLETLKLAERVADGD 324
```

RESULT 13
Q8IJZ2
ID Q8IJZ2 PRELIMINARY; PRT; 202 AA.
AC Q8IJZ2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF10_0048.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014829; AAN35246.1; -.
KW Hypothetical protein.
SQ SEQUENCE 202 AA; 24259 MW; 48FD22A7F21165A6 CRC64;

Query Match 7.7%; Score 99; DB 5; Length 202;
Best Local Similarity 24.9%; Pred. No. 8.9;
Matches 46; Conservative 31; Mismatches 50; Indels 58; Gaps 9;

```
Qy           4 KYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHN-----EKL--TSSIEYKRHGDDYAK 56
Db          32 KYEENIILLKKLMRTENIKF----------DQKYNAINILEKLKDIEDIQYEDILNDYAK 81

Qy          57 --------YAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKK------ 102
Db          82 NFLTFLKGSFEKVAEEIQI---NDKISFRRNE--LIKQDLNNKLQNLNHNYNLEYINLNK 136

Qy         103 ---TE------TTLIEQNMLSKILERSLEEMDDEEVK-------------EMCDELSIKN 140
Db         137 CLGTEDFESQLNASLKFNSLKDILSKLLKEMKDYEIQYNNLSDTDKFLDMKICELENYLN 196

Qy         141 TDNLN 145
Db         197 NTNQN 201
```

RESULT 14
O43633
ID O43633 PRELIMINARY; PRT; 222 AA.
AC O43633;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BC-2 protein (Putative breast adenocarcinoma marker) (32KD) (32
DE kDa).
GN BC-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```
RA   Slater C., Thill G., Obar R.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Koczan D., Reimer T., Rump A., Merck-Rousseau M.F., Rosenthal A.,
RA   Friese K., Thiesen H.J.;
RT   "Role of the BC-2 gene in breast cancer (in preparation).";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Skin;
RA   Strausberg R.;
RL   Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Kalnine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RA   Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,
RA   Phelan M., Farmer A.;
RT   "Cloning of human full-length CDSs in BD Creator(TM) System Donor
RT   vector.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF042384; AAC00005.1; -.
DR   EMBL; AJ277113; CAC14310.1; -.
DR   EMBL; BC002502; AAH02502.1; -.
DR   EMBL; BC047005; AAH47005.1; -.
DR   EMBL; BT007298; AAP35962.1; -.
DR   InterPro; IPR005024; DUF279.
DR   Pfam; PF03357; SNF7; 1.
SQ   SEQUENCE   222 AA;  25104 MW;  F2B86C623829E32E CRC64;

  Query Match             7.7%;  Score 99;  DB 4;  Length 222;
  Best Local Similarity   21.4%;  Pred. No. 10;
  Matches   39;  Conservative   48;  Mismatches   83;  Indels   12;  Gaps    5;

Qy          6 DRDLEFLKQLESSDLLDLFEVLVFGK-DGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEE 64
              ||: : |: |   : |: :: |: | :  : | : | |          : :: :
Db         30 DRERQKLETQEKKIIADIKKMAKQGQMDAVRIMAKDLVRTRRYVRKFVLMRANIQAVSLK 89

Qy         65 LQYYGS-NSFASFIKG----EGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119
               :|   | || |  :||     | : :::      |: : : :: |   ::: |::  ::
Db         90 IQTLKSNNSMAQAMKGVTKAMGTMNRQLKLPQIQKIMMEFERQAEIMDMKEEMMNDAIDD 149

Qy        120 SLEEMDDEE-----VKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANA 174
              :: : :|||     | :: ||| :  || |:    : :|::   || |:   |  :|:|
Db        150 AMGDEEDEEESDAVVSQVLDELGLSLTDELSNLPSTGGSLSV-AAGGKKAEAAASALADA 208

Qy        175 VA 176
               |
Db        209 DA 210


RESULT 15
Q9DB34
ID   Q9DB34      PRELIMINARY;      PRT;   222 AA.
AC   Q9DB34;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   1500016L11Rik protein (RIKEN cDNA 1500016L11 gene).
GN   1500016L11RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast tumor;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK005267; BAB23919.1; -.
DR   EMBL; BC012230; AAH12230.1; -.
DR   MGD; MGI:1916203; 1500016L11Rik.
DR   InterPro; IPR005024; DUF279.
DR   Pfam; PF03357; SNF7; 1.
SQ   SEQUENCE   222 AA;  25134 MW;  F2B86C623832E29E CRC64;

  Query Match             7.7%;  Score 99;  DB 11;  Length 222;
  Best Local Similarity   21.4%;  Pred. No. 10;
  Matches   39;  Conservative   48;  Mismatches   83;  Indels   12;  Gaps    5;

Qy          6 DRDLEFLKQLESSDLLDLFEVLVFGK-DGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEE 64
              ||: : |: |   : |: :: |: | :  : | : | |          : :: :
Db         30 DRERQKLETQEKKIIADIKKMAKQGQMDAVRIMAKDLVRTRRYVRKFVLMRANIQAVSLK 89

Qy         65 LQYYGS-NSFASFIKG----EGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119
               :|   | || |  :||     | : :::      |: : : :: |   ::: |::  ::
Db         90 IQTLKSNNSMAQAMKGVTKAMGTMNRQLKLPQIQKIMMEFERQAEIMDMKEEMMNDAIDD 149

Qy        120 SLEEMDDEE-----VKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANA 174
              :: : :|||     | :: ||| :  || |:    : :|::   || |:   |  :|:|
Db        150 AMGDEEDEEESDAVVSQVLDELGLSLTDELSNLPSTGGSLSV-AAGGKKAEATASALADA 208

Qy        175 VA 176
               |
Db        209 DA 210


Search completed: July  5, 2004, 03:12:04
Job time : 29 secs
```

GenCore version 5.1.6

CM protein - protein search, using sw model

Run on: July 5, 2004, 03:07:28 ; Search time 16 Seconds
(without alignments)
816.336 Million cell updates/sec

Title: US-09-732-091-4
Perfect score: 1279
Sequence: 1 MAYKYDRDLEFLKQLESSDL.....LRLKTQQANGDKKSLQIESI 253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	364	28.5	273	4	US-09-252-991A-29841	Sequence 29841, A
2	346.5	27.1	321	4	US-09-252-991A-29840	Sequence 29840, A
3	247	19.3	258	4	US-09-489-039A-8852	Sequence 8852, Ap
4	100	7.8	739	4	US-09-543-681A-6437	Sequence 6437, Ap
5	99	7.7	222	3	US-08-944-604-16	Sequence 16, Appl
6	99	7.7	274	4	US-09-134-001C-5279	Sequence 5279, Ap
7	96	7.5	173	2	US-08-658-639-12	Sequence 12, Appl
8	96	7.5	173	3	US-08-944-604-12	Sequence 12, Appl
9	95	7.4	2125	4	US-09-919-172-29	Sequence 29, Appl
10	91.5	7.2	1086	4	US-09-543-681A-7696	Sequence 7696, Ap
11	89.5	7.0	281	4	US-09-198-452A-749	Sequence 749, App
12	89.5	7.0	543	4	US-09-328-352-5845	Sequence 5845, Ap
13	89.5	7.0	546	4	US-09-345-236B-98	Sequence 98, Appl
14	89.5	7.0	546	4	US-09-345-236B-121	Sequence 121, App
15	89.5	7.0	560	4	US-09-446-301A-50	Sequence 50, Appl
16	88.5	6.9	552	4	US-09-446-301A-4	Sequence 4, Appli
17	88.5	6.9	552	4	US-09-099-932-4	Sequence 4, Appli
18	88	6.9	488	4	US-09-489-039A-13363	Sequence 13363, A
19	87.5	6.8	732	2	US-08-533-669A-18	Sequence 18, Appl
20	87.5	6.8	732	4	US-09-307-143-4	Sequence 4, Appli
21	87.5	6.8	732	4	US-09-183-861-18	Sequence 18, Appl
22	87.5	6.8	732	4	US-09-022-765-18	Sequence 18, Appl
23	87.5	6.8	732	4	US-09-551-974A-18	Sequence 18, Appl
24	87.5	6.8	732	4	US-09-565-501A-18	Sequence 18, Appl
25	87.5	6.8	732	4	US-09-639-206A-18	Sequence 18, Appl
26	87.5	6.8	732	4	US-09-874-923-18	Sequence 18, Appl
27	87	6.8	491	3	US-09-029-267-2	Sequence 2, Appli
28	86.5	6.8	331	4	US-09-134-001C-3626	Sequence 3626, Ap
29	84.5	6.6	528	4	US-08-637-670-27	Sequence 27, Appl
30	84.5	6.6	691	4	US-09-830-217-12	Sequence 12, Appl
31	84	6.6	492	4	US-09-107-532A-7062	Sequence 7062, Ap
32	83.5	6.5	264	2	US-08-719-697-8	Sequence 8, Appli
33	83.5	6.5	264	3	US-08-727-616A-8	Sequence 8, Appli
34	83.5	6.5	264	4	US-09-481-756-8	Sequence 8, Appli
35	83.5	6.5	286	1	US-07-721-775A-2	Sequence 2, Appli
36	83.5	6.5	286	1	US-08-339-658-2	Sequence 2, Appli
37	83.5	6.5	286	3	US-09-263-933-7	Sequence 7, Appli
38	83.5	6.5	286	3	US-09-263-933-14	Sequence 14, Appl
39	83.5	6.5	286	3	US-09-263-933-21	Sequence 21, Appl
40	83.5	6.5	286	4	US-09-025-769B-265	Sequence 265, App
41	83.5	6.5	286	4	US-09-025-769B-362	Sequence 362, App
42	83.5	6.5	286	4	US-09-919-901-7	Sequence 7, Appli
43	83.5	6.5	286	4	US-09-919-901-14	Sequence 14, Appl
44	83.5	6.5	286	4	US-09-919-901-21	Sequence 21, Appl
45	83.5	6.5	299	4	US-09-025-769B-285	Sequence 285, App

ALIGNMENTS

RESULT 1
US-09-252-991A-29841
; Sequence 29841, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29841
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29841

Query Match 28.5%; Score 364; DB 4; Length 273;
Best Local Similarity 35.4%; Pred. No. 1.1e-30;
Matches 85; Conservative 49; Mismatches 88; Indels 18; Gaps 4;

```
Qy          10 EFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEELQYYG 69
               : |::: : ||  | | ::      |   | |:  :::||   :: :||  | :||: :|
Db          28 QLLERVSNDDLEPLVEYIL------KARTESLSKQVDFKRWHPEHRRYASAILDELRLPG 81

Qy          70 SNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSLEEMDDEEV 129
               ||||:  :  |  | |:: ||  ||||      |   :|: |:  || ::||:   |:
Db          82 GNSFANLWRKSGPSYTEVVRDVAGKLKVKGVGSMELIELEEAMVQSILRQALEKSSGEDR 141

Qy         130 KEMCDEL------SIKNTDNLNRQALSAATLTLFKMGGPKSYQLAVIVANAVAKTILGRG 183
               :|: : |      | |  ||  |||   | :  :     |: : :: |::|: :|| |
Db         142 RELEEILREAGLDKTKMTALLNGSALSG--LVVPAVARMILYRTSTVIVNSMAQQLLGHG 199

Qy         184 LSLA----GNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQ 239
               |  |    |   | :: | |||||:| |||||:|:||||||||||  : :| |||| :
Db         200 LRSAVVAGGTFAGGRAVAALAGPVGWVIAGVWTAVDLAGPAYRVTIPCVLHIAMLRLKAR 259
```

RESULT 2
US-09-252-991A-29840
; Sequence 29840, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

```
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 29840
;   LENGTH: 321
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29840

  Query Match             27.1%;  Score 346.5;  DB 4;  Length 321;
  Best Local Similarity   34.3%;  Pred. No. 1.1e-28;
  Matches   82;  Conservative   42;  Mismatches  112;  Indels    3;  Gaps    3;

Qy          1 MAYKYDRDL-EFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAE 59
              ||  :| || : |    | |:  | :|:    :|    :  :   :   : |
Db         53 MAIIHDADLADVLLSASSDDIRLLIDVITDNGNGRISLSSSVCRQLSAAKEGVVGEFERG 112

Qy         60 RIAEELQYYGSNSFASFIK-GEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILE 118
               :||||  :| ||  :  : | || |||:| ||    : |:     :   :|  :::|: |
Db        113 MVAEELMRFGGNSLMNLFRGGSGVPYKELLSDVASHVGVSKESTGDCARMEMAIITKVFE 172

Qy        119 RSLEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKT 178
              :|:  | :|:     : |       :   || || :      |  || || :||:|   :
Db        173 QSIGRMSEEDKATFFESLGTSYRSGMGPVAL-AALIASLSASGMTSYGLAAMVASATMSS 231

Qy        179 ILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLK 237
              ::|||::|||   | | || ||||||| | |:||| |:| |||||:|  | :  :| |
Db        232 LVGRGVALAGGATLGRGLSVLTGPVGWAIAGIWTAFDLASPAYRVTLPCVIQIGHMRQK 290


RESULT 3
US-09-489-039A-8852
; Sequence 8852, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 8852
;   LENGTH: 258
;   TYPE: PRT
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8852

  Query Match             19.3%;  Score 247;  DB 4;  Length 258;
  Best Local Similarity   28.2%;  Pred. No. 3.4e-18;
  Matches   70;  Conservative   48;  Mismatches  100;  Indels   30;  Gaps    6;

Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSS---IEYKRHGDDYAKYAE 59
              | | || || :   :  |      :|  |::|: | :  |  :   :  : | : | :  :
Db         14 YLEDDDLAFLPECSEAHLEAFTRILTHGENGKPRLSSTLLRNETFLAMEGHPERYRRNWQ 73

Qy         60 RIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119
               || :||::| :| |: ::   |  |: || ||| :||   :|:  |  ||| :|:  |:
Db         74 LIAGELQHFGGDSIANTLRRHGKFYRAILLDVCKRLKAKVDKQLSTPQIEQQLLAHFLQH 133

Qy        120 SLEEMDDEEVKEMCDELSIKN--TDNL------NRQALSAATLTLFKMGGFKSYQLAVIV 171
              |  ::: |:  :    :  ::   |:|      :|:      :| |          :|  |:
Db        134 SWNKLNAEQKAQFLAAVECRSHELDSLMAHLLRHRKLSEGVTLLL-------DERLTAIL 186

Qy        172 ANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVV 231
                | :::| ||                | ||:|  :  |     ::| |||||||| : :
Db        187 RTHAAVSVIGHGLVRGAG---------LGGPLGAALNSVKA---VSGSAYRVTIPAVLHI 234

Qy        232 ATLRLKTQ 239
              | ||   |
Db        235 ACLRQMLQ 242


RESULT 4
US-09-543-681A-6437
; Sequence 6437, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6437
;   LENGTH: 739
;   TYPE: PRT
;   ORGANISM: Proteus mirabilis
US-09-543-681A-6437

  Query Match              7.8%;  Score 100;  DB 4;  Length 739;
  Best Local Similarity   25.6%;  Pred. No. 0.089;
  Matches   54;  Conservative   42;  Mismatches   65;  Indels   50;  Gaps   10;

Qy          6 DRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEEL 65
              |:  :||  | :| |  | |:   | |  |:  |:: :||:      | | | :|   :
Db        517 DKVSDFLVSLVNSSLDGLKEL---GLD--KKIIEEMKNSIQ------DKLKMATKILVTV 565

Qy         66 QYYGSNSFASFIKGEGVLYKEILCDVCDK---------LKVNYNKKTETTL--IEQNMLS 114
              : : :  ||: |  :   : : |  :|          ||    |  | |  : :::::
Db        566 VLFVAATALSFVIGPAM---KGISDAVNKISNQQIRQILKKVLNDGLEAVLGKMIKDIII 622

Qy        115 KILERSLEEMDDEEVKEMCDELSIKNTDNLNRQAL------SAATLTLFKMGGFKSYQLA 168
              | || :|::|    |::  |:| | :  |||  :       |||| |:   |
Db        623 KALEEALEKID----KQLAKEISKKASIMLNRTVVASKLTNSAATNTVNIYGS------- 671

Qy        169 VIVANAVAKTILGRGLSLAGNQVLTRTLSFL 199
               ::|: : :       |:||:: ||  |  :
Db        672 -VIASKIIQ-------SIAGSKKLTAVLDII 694


RESULT 5
US-08-944-604-16
; Sequence 16, Application US/08944604
; Patent No. 6218131
;  GENERAL INFORMATION:
;    APPLICANT:  KEESEE, SUSAN
;    APPLICANT:  OBAR, ROBERT
;    APPLICANT:  WU, YING-JYE
;    TITLE OF INVENTION:  MATERIALS AND METHODS FOR DETECTION OF
;    TITLE OF INVENTION:  BREAST CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Testa, Hurwitz & Thibeault
;      STREET:  125 High St.
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/944,604
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, THOMAS C
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  MTP-021 (8395/24)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  222 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-944-604-16

  Query Match             7.7%;  Score 99;  DB 3;  Length 222;
  Best Local Similarity   21.4%;  Pred. No. 0.018;
  Matches   39;  Conservative   48;  Mismatches   83;  Indels   12;  Gaps    5;

Qy          6 DRDLEFLKQLESSDLLDLFEVLVFGK-DGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEE 64
              ||: : |:  |   : |: ::   |: |  :   : |  :  | |           : :: :
Db         30 DRERQKLETQEKKIIADIKKMAKQGQMDAVRIMAKDLVRTRRYVRKFVLMRANIQAVSLK 89

Qy         65 LQYYGS-NSFASFIKG----EGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119
                :|   | || |  :||      | : :::      |: : : :: |   ::: |::  ::
Db         90 IQTLKSNNSMAQAMKGVTKAMGTMNRQLKLPQIQKIMMEFERQAEIMDMKEEMMNDAIDD 149

Qy        120 SLEEMDDEE-----VKEMCDELSIKNTDNLNRQALSAATLTLFKMGGPKSYQLAVIVANA 174
              :: : :|||     | :: ||| :  || |:    :  :|::   || |:   |  :|:|
Db        150 AMGDEEDEEESDAVVSQVLDELGLSLTDELSNLPSTGGSLSV-AAGGKKAEAAASALADA 208

Qy        175 VA 176
              |
Db        209 DA 210


RESULT 6
US-09-134-001C-5279
; Sequence 5279, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5279
;   LENGTH: 274
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5279

  Query Match             7.7%;  Score 99;  DB 4;  Length 274;
  Best Local Similarity   23.8%;  Pred. No. 0.025;
  Matches   62;  Conservative   44;  Mismatches   88;  Indels   66;  Gaps   14;

Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSS----------------IE 46
              | |: ||                  |:  | |:: : : : :|                :|
Db          4 YNYNNDL----------------FLILKKFGDEMNKDTIETSREALIKDVVMIAARILLE 47

Qy         47 YKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKT-ET 105
                    |       |||  :| |  ||||  :     | ||: |     :        |  || :|
Db         48 SGAEGTRVEDTMARIATKLGYPESNSFVTNTVIEFVLHNEAYPRL-------YRIKTRDT 100

Qy        106 TLIEQNMLSKILERSLE-EMDDEEVKEMCDELSI-KNTDNLNRQALSA---ATLTLFKMG 160
              ||: :  ::|  :      |  || |   :|: : |   :|  : ::|    ||  |:  |
Db        101 NLIKISQANEISRQITNGTMTLEEAKYQLEEIYVAKRDSSLPFKGIAAAIIATSFLYLQG 160

Qy        161 GFKSYQLAVIVANAVAKTILGRGLSLAGNQVLTRTL--SFLTGPVGWIITGVWTAIDIAG 218
              |    :|  |:   :| ||   |  :   ::| | |   |:   :| :: |:   | : |
Db        161 G----RLVDIITAVLAGTI---GYLVV--EILDRKLHAQFIPEFIGSLVIGI---ISVIG 208

Qy        219 PAYRVTIP----ACIVVATL 234
              |:   :|    | |::| :
Db        209 HAF---VPSGDLATIIIAAV 225


RESULT 7
US-08-658-639-12
; Sequence 12, Application US/08658639
; Patent No. 5914238
;  GENERAL INFORMATION:
;    APPLICANT:  KEESEE, SUSAN
;    APPLICANT:  OBAR, ROBERT
;    APPLICANT:  WU, YING-JYE
;    TITLE OF INVENTION:  MATERIALS AND METHODS FOR DETECTION OF
;    TITLE OF INVENTION:  BREAST CANCER
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Testa, Hurwitz & Thibeault
;      STREET:  125 High St.
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,639
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, THOMAS C
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  MTP-021 (8395/24)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  173 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-658-639-12

  Query Match             7.5%;  Score 96;  DB 2;  Length 173;
  Best Local Similarity   23.3%;  Pred. No. 0.027;
  Matches   27;  Conservative   33;  Mismatches   46;  Indels   10;  Gaps    3;

Qy         70 SNSFASFIKG----EGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSLEEMD 125
              :|| |  :||      | : :::      |: : : :: |   ::: |::  :: :: : :
Db         47 NNSMAQAMKGVTKAMGTMNRQLKLPQIQKIMMEFERQAEIMDMKEEMMNDAIDDAMGDEE 106

Qy        126 DEE-----VKEMCDELSIKNTDNLNRQALSAATLTLFKMGGPKSYQLAVIVANAVA 176
              |||     | :: ||| :  || |:    :  :|::   || |:   |  :|:| |
Db        107 DEEESDAVVSQVLDELGLSLTDELSNLPSTGGSLSV-AAGGKKAEAAASALADADA 161


RESULT 8
US-08-944-604-12
; Sequence 12, Application US/08944604
; Patent No. 6218131
;  GENERAL INFORMATION:
```

; APPLICANT: KEESEE, SUSAN
; APPLICANT: OBAR, ROBERT
; APPLICANT: WU, YING-JYE
; TITLE OF INVENTION: MATERIALS AND METHODS FOR DETECTION OF
; TITLE OF INVENTION: BREAST CANCER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Testa, Hurwitz & Thibeault
; STREET: 125 High St.
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,604
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MEYERS, THOMAS C
; REGISTRATION NUMBER: 36,989
; REFERENCE/DOCKET NUMBER: MTP-021 (8395/24)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 173 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-944-604-12

Query Match 7.5%; Score 96; DB 3; Length 173;
Best Local Similarity 23.3%; Pred. No. 0.027;
Matches 27; Conservative 33; Mismatches 46; Indels 10; Gaps 3;

```
Qy          70 SNSFASPIKG----EGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSLEEMD 125
               :|| |  :||     | : :::      |: : : :: |   ::: |::  :: :: : :
Db          47 NNSMAQAMKGVTKAMGTMNRQLKLPQIQKIMMEFERQAEIMDMKEEMMNDAIDDAMGDEE 106

Qy         126 DEE-----VKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVA 176
                   |||     | :: ||| :  || |:    :  :|::  || |:   |  :|:| |
Db         107 DEEESDAVVSQVLDELGLSLTDELSNLPSTGGSLSV-AAGGKKAEAAASALADADA 161
```

RESULT 9
US-09-919-172-29
; Sequence 29, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 2125
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673545 3774181CD1
US-09-919-172-29

Query Match 7.4%; Score 95; DB 4; Length 2125;
Best Local Similarity 25.4%; Pred. No. 1.5;
Matches 46; Conservative 35; Mismatches 62; Indels 38; Gaps 9;

```
Qy           2 AYKYDRDLEFLK--QLESSDLLDLF----EVLVFGKDGEKRHNEKLTSSIEYKRHGDDYA 55
               ::: :::|| |:  | :|  | : |    | |:     || :|:|:    |     :: |
Db         986 SFRDEKELERLQICQRKSDHLKEQFEKSHEQLLQNIKAEKENNDKIQRLNEELEKSNECA 1045

Qy          56 KYAERIAEEL--QYYGSNSFASFIKGEG---VLYKEILCDVCDKLKV------------- 97
               :  ::  |||  |   :       |: |     || |: :   |: ||:
Db        1046 EMLKQKVEELTRQNNETKLMMQRIQAESENIVLEKQTIQQRCEALKIQADGFKDQLRSTN 1105

Qy          98 -NYNKKTETTLIEQNMLSKI--LERSLEEMDD--EEVKEMCDELSI------KNTDNLNR 146
                : :|:|:|   ||:   ||  ||  | :  :   | |: ||: :|     |   |||
Db        1106 EHLHKQTKT---EQDFQRKIKCLEEDLAKSQNLVSEFKQKCDQQNIIIQNTKKEVRNLNA 1162

Qy         147 Q 147
               :
Db        1163 E 1163
```

RESULT 10
US-09-543-681A-7696
; Sequence 7696, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7696
; LENGTH: 1086
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7696

Query Match 7.2%; Score 91.5; DB 4; Length 1086;
Best Local Similarity 19.5%; Pred. No. 1.3;
Matches 43; Conservative 38; Mismatches 73; Indels 67; Gaps 9;

```
Qy           3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAERIA 62
               : |||  | ::|::          ||  : ||   :   |||  :
Db         786 FAYDRGKE-IEQMQ-------FETTLIGK--SRAEQEKLNA------------------ 816

Qy          63 EELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSLE 122
                          ::   |||::   |: :|  ||  :  | |  :| :  :: :|
Db         817 --------------LRQIDVLYQQASVDLGEKELVNLQRNVELT--KQQIEEELRKREAM 860

Qy         123 EMD------------DEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVI 170
               : |              |   :: : :    |: ||  : : |   |   | || : |||
Db         861 KGDPMAGLKQGLSDFSESAMDVMENVRNVTTNALNNMSDALADFALTGKGSFKDFANAVI 920

Qy         171 --VANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITG 209
                 :   | | :: : :   |     : : |    :||:  |
Db         921 SDITRMVMKMLIFKAIEAGG-----QAMGF---DMGWMSKG 953
```

RESULT 11
US-09-198-452A-749
; Sequence 749, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragme
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, pr
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999

; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 749
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-749

Query Match 7.0%; Score 89.5; DB 4; Length 281;
Best Local Similarity 22.0%; Pred. No. 0.27;
Matches 56; Conservative 49; Mismatches 103; Indels 47; Gaps 11;

Qy 22 DLFEVLVFGKDGEKRHNEKLTSS-----IEYKRHGDDYAKYAERIAE--------ELQYY 68
Db 30 DSRDVKVFSLEGKQTRQEKTTSSKGNTRTESRKFADEEKRVDDEIAEVGSKEEEQESQEF 89

Qy 69 --GSNSFA--SFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSLEEM 124
Db 90 CLAENAFAGMSLIDIAAAGSAEAVVEVA-PIAVS---SIDTQWIENIILSTVESMVISEI 145

Qy 125 DDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGL 184
Db 146 NGEQLVEL-----VLDASSSVPEAFVGANLTLVQSG----QDLSVKFSSFVDATQMAEAA 196

Qy 185 SLAGN--QVLTRTLSFLTG--------PVGWIITGVWTAIDIAGPAYRVTIPACIVVATL 234
Db 197 DLVTNNPSQLSSLVSALKGHQLTLKEFSVGNLLVQLPKIEEVQTPLH-------MIASTI 249

Qy 235 RLKTQQANGDKKSLQ 249
Db 250 RHREEKDQRDQNQKQ 264

RESULT 12
US-09-328-352-5845
; Sequence 5845, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5845
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5845

Query Match 7.0%; Score 89.5; DB 4; Length 543;
Best Local Similarity 21.1%; Pred. No. 0.74;
Matches 40; Conservative 31; Mismatches 52; Indels 67; Gaps 10;

Qy 6 DRDLEFLKQLESSDLLDLPEVLVFGKDGEKRHNEKLTSSIEYKRHGD----------DYA 55
Db 366 DKSLEELEKEERKDLKNRFD-------------------EYKNKGQIKFVTFHQSFSYE 405

Qy 56 KYAERI-AEELQYYGSNSFASFIKGEGVLYKEILCD------VCDKLKVNYNKKTETTLI 108
Db 406 DFVEGIRAETVEYSDGKKNIEYPVVSGVF--KLLCDTAQSKVILESQKINFDSNT----- 458

Qy 109 EQNMLSKILERSLEEM-DDEEVKEMC-----------DELSIKNTDNLNRQALSAATLTL 156
Db 459 -----NEIWKMSLGRAGEDEDIFDYCIKNHCVLLGWGDELDFSNA--VNRKQIE----EI 507

Qy 157 FKMGGFKSYQ 166
Db 508 MDKNGYEAYR 517

RESULT 13
US-09-345-236B-98
; Sequence 98, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
; APPLICANT: Becnel, James J.
; APPLICANT: Tukuo, Fukuda
; APPLICANT: Moser, Bettina
; APPLICANT: Cockburn, Andrew
; APPLICANT: White, Susan E.
; APPLICANT: Undeen, Albert H.
; TITLE OF INVENTION: No. 6521454el Baculoviruses, Insecticidal
; TITLE OF INVENTION: Compositions, and Methods for Control of Invertebrates
; FILE REFERENCE: 21042.0004
; CURRENT APPLICATION NUMBER: US/09/345,236B
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 546
; TYPE: PRT
; ORGANISM: mosquito baculovirus
US-09-345-236B-98

Query Match 7.0%; Score 89.5; DB 4; Length 546;
Best Local Similarity 23.5%; Pred. No. 0.75;
Matches 60; Conservative 34; Mismatches 92; Indels 69; Gaps 10;

Qy 23 LFEVLVFGKDGEKRHNEK-------LTSSIEYKRHGDDYAKYAE---RIAEELQYYGSNS 72
Db 153 VFEVVGGGPEGDYAAGEEDEVSRNSLNFDMASEVQSTDAAKVMELFSALSEEQRNVILNN 212

Qy 73 FASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSLEEMDDEEVKEM 132
Db 213 FGAAPSGSGT--------------------TPPTSAQPDM----------EVEDVETVEK 242

Qy 133 CDELSIKNTDNLN----RQALSAATLTLFKMG---GFKSYQLAVIVANAVAKTILGRGLS 185
Db 243 PENLNDIITDQLRDFMAQELKKAAENYVPKWGSTVGESKSALAITVADRVSRSFMYEGRI 302

Qy 186 LAGNQV-----------LTRTLSFLTGPVGWII-------TGVWTAIDIAGPAYRVTIPA 227
Db 303 VDYNQVVLHILDNYDQRLEELLSFRTKT--YIIAEGVPHDSKVHDYVDLT--QYRETVPY 358

Qy 228 CIVVATLRLKTQQAN 242
Db 359 SIALNNLSRGVDQAN 373

RESULT 14
US-09-345-236B-121
; Sequence 121, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
; APPLICANT: Becnel, James J.
; APPLICANT: Tukuo, Fukuda
; APPLICANT: Moser, Bettina
; APPLICANT: Cockburn, Andrew
; APPLICANT: White, Susan E.
; APPLICANT: Undeen, Albert H.
; TITLE OF INVENTION: No. 6521454el Baculoviruses, Insecticidal
; TITLE OF INVENTION: Compositions, and Methods for Control of Invertebrates
; FILE REFERENCE: 21042.0004
; CURRENT APPLICATION NUMBER: US/09/345,236B
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 121
; LENGTH: 546
; TYPE: PRT
; ORGANISM: mosquito baculovirus
US-09-345-236B-121

```
  Query Match             7.0%;  Score 89.5;  DB 4;  Length 546;
  Best Local Similarity   23.5%;  Pred. No. 0.75;
  Matches   60;  Conservative   34;  Mismatches   92;  Indels   69;  Gaps   10;

Qy         23 LFEVLVFGKDGEKRHNEK-------LTSSIEYKRHGDDYAKYAE---RIAEELQYYGSNS 72
              :|||:  | :|:    |:       |   :  :    | ||  |    ::|| :    |:
Db        153 VFEVVGGGPEGDYAAGEEDEVSRNSLNFDMASEVQSTDAAKVMELFNALSEEQRNVILNN 212

Qy         73 FASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERSLEEMDDEEVKEM 132
              |  :   | |                        |  |  : :|          |::| |  |
Db        213 FGAAPSGSGT-------------------TPPTSAQPDM----------EVEDVETVEK 242

Qy        133 CDELSIKNTDNLN----RQALSAATLTLFKMG---GFKSYQLAVIVANAVAKTILGRGLS 185
               : |:   || |     ::   ||   : | |   |     ||: ||: |::: :  |
Db        243 PENLNDIITDQLRDFMAQELKKAAENYVPKWGSTVGESKSALAITVADRVSRSFMYEGRI 302

Qy        186 LAGNQV----------LTRTLSPLTGPVGWII-------TGVWTAIDIAGPAYRVTIPA 227
              :  |||           |   |||  |    :||       : |   :|:    || |:|
Db        303 VDYNQVVLHILDNYDQRLEELLSPRTKT--YIIAEGVPHDSKVHDYVDLT--QYRETVPY 358

Qy        228 CIVVATLRLKTQQAN 242
              | :  |     |||
Db        359 SIALNNLSRGVDQAN 373


RESULT 15
US-09-446-301A-50
; Sequence 50, Application US/09446301A
; Patent No. 6506893
; GENERAL INFORMATION:
;  APPLICANT: EL SOLH, NEVINE
;  APPLICANT:  ALLIGNET, JEANINE
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND THEIR USE FOR DETECTING RESISTANCE
;  TITLE OF INVENTION:  TO STREPTOGRAMIN A OR TO STREPTOGRAMIN B AND RELATED
;  TITLE OF INVENTION:  COMPOUNDS
;  FILE REFERENCE: 03715-0059
;  CURRENT APPLICATION NUMBER: US/09/446,301A
;  CURRENT FILING DATE:  1999-12-20
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
;   LENGTH: 560
;   TYPE: PRT
;   ORGANISM: Staphylococcus sp.
US-09-446-301A-50

  Query Match             7.0%;  Score 89.5;  DB 4;  Length 560;
  Best Local Similarity   25.3%;  Pred. No. 0.78;
  Matches   37;  Conservative   31;  Mismatches   47;  Indels   31;  Gaps    7;

Qy          8 DLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEELQY 67
              |:: :: ||  :|| :| :|     :|                      |: : :||:|
Db        435 DIDAVEALE--ELLITYEGVVLFASHDK-------------------KFIQNLAEQLLI 472

Qy         68 YGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKK--TETTLIEQNMLSKILERSLEEMD 125
                :|    |  ||  | | |  : || |:| |:|   |  :| :  :| :| :   | :
Db        473 IENNKVKKF---EGT-YIEYL-KIKDKPKLNTNEKELKEKKMILEMQISSLLSKISMEEN 527

Qy        126 DEEVKEMCD--ELSIKNTDNLNRQAL 149
              :|: ||: :  :| :|   :||:  :
Db        528 EEKNKELDEKYKLKLKELKSLNKNIM 553


Search completed: July  5, 2004, 03:12:30
Job time : 17 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 5, 2004, 03:06:13 ; Search time 39 Seconds
(without alignments)
1832.936 Million cell updates/sec

Title: US-09-732-091-4
Perfect score: 1279
Sequence: 1 MAYKYDRDLEFLKQLESSDL.....LRLKTQQANGDKKSLQIESI 253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1279	100.0	253	5	AAE26860	Aae26860 Helicobac
2	1279	100.0	265	5	AAE26878	Aae26878 Helicobac
3	1270	99.3	253	2	AAW20486	Aaw20486 H. pylori
4	1270	99.3	253	2	AAW24673	Aaw24673 H. pylori
5	1270	99.3	256	2	AAW20866	Aaw20866 H. pylori
6	1264	98.8	253	5	AAE26880	Aae26880 Helicobac
7	1264	98.8	265	5	AAE26879	Aae26879 Helicobac
8	472.5	36.9	155	2	AAW98774	Aaw98774 H. pylori
9	252	19.7	49	5	AAE26876	Aae26876 Helicobac
10	198	15.5	38	5	AAE26873	Aae26873 Helicobac
11	194	15.2	41	5	AAE26875	Aae26875 Helicobac
12	152	11.9	30	5	AAE26874	Aae26874 Helicobac
13	150	11.7	30	5	AAE26872	Aae26872 Helicobac
14	99	7.7	222	4	AAB98720	Aab98720 Human bre
15	99	7.7	274	5	ABP40434	Abp40434 Staphyloc
16	98.5	7.7	1009	6	ABU15908	Abu15908 Protein e
17	98	7.7	413	6	ABM71523	Abm71523 Staphyloc
18	97	7.6	916	6	ABU48566	Abu48566 Protein e
19	96.5	7.5	701	5	ABP73563	Abp73563 Candida a
20	96	7.5	173	2	AAW44365	Aaw44365 Breast ca
21	96	7.5	173	4	AAB98717	Aab98717 Human bre
22	95	7.4	517	6	ABU26619	Abu26619 Protein e
23	95	7.4	2125	5	ABG77172	Abg77172 Prostate
24	95	7.4	2649	7	ADB79929	Adb79929 Human put
25	93.5	7.3	1196	6	ABU24813	Abu24813 Protein e
26	93	7.3	381	4	AAG82981	Aag82981 S. epider
27	93	7.3	610	6	ABU19293	Abu19293 Protein e
28	92.5	7.2	996	4	AAU33755	Aau33755 Staphyloc
29	92.5	7.2	1009	4	AAU36548	Aau36548 Staphyloc
30	92	7.2	309	5	ABP30227	Abp30227 Streptoco
31	92	7.2	327	5	ABP28162	Abp28162 Streptoco
32	91.5	7.2	187	3	AAG44550	Aag44550 Arabidops
33	91.5	7.2	187	3	AAG48726	Aag48726 Arabidops
34	91.5	7.2	241	3	AAG44549	Aag44549 Arabidops
35	91.5	7.2	241	3	AAG48725	Aag48725 Arabidops
36	91.5	7.2	256	3	AAG48724	Aag48724 Arabidops
37	91.5	7.2	256	3	AAG44548	Aag44548 Arabidops
38	91.5	7.2	375	5	ABP30479	Abp30479 Streptoco
39	91.5	7.2	390	5	ABP27884	Abp27884 Streptoco
40	91.5	7.2	1769	6	ABR52974	Abr52974 Protein s
41	91.5	7.2	1847	3	AAY51631	Aay51631 M. jannas
42	91.5	7.2	1847	3	AAY52002	Aay52002 M. jannas
43	90.5	7.1	1411	2	AAY31948	Aay31948 Plasmodiu
44	90.5	7.1	1417	2	AAY31947	Aay31947 Plasmodiu
45	90	7.0	216	4	AAG82765	Aag82765 S. epider

ALIGNMENTS

RESULT 1
AAE26860
ID AAE26860 standard; protein; 253 AA.
XX
AC AAE26860;
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30 protein.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44513.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Claim 1; Page 111-112; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is

```
CC   Helicobacter sp. HP30 protein
XX
SQ   Sequence 253 AA;

  Query Match           100.0%;  Score 1279;  DB 5;  Length 253;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-118;
  Matches  253;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              |||||||||||||
Db        241 ANGDKKSLQIESI 253


RESULT 2
AAE26878
ID   AAE26878 standard; protein; 265 AA.
XX
AC   AAE26878;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Helicobacter sp. PQE/HP30 protein.
XX
KW   HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW   type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW   low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW   immunostimulant; cytostatic; peptic ulcer.
XX
OS   Helicobacter sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 12
FT                   /note= "Encoded by TTC"
XX
PN   WO200251237-A2.
XX
PD   04-JUL-2002.
XX
PF   07-DEC-2001; 2001WO-US048392.
XX
PR   07-DEC-2000; 2000US-00732091.
XX
PA   (ANTE-) ANTEX BIOLOGICS INC.
XX
PI   Tian J,  Walker R,  Jackson WJ;
XX
DR   WPI; 2002-666854/71.
DR   N-PSDB; AAD44535.
XX
PT   Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT   the proteins, useful as vaccines for raising immune response in animals.
XX
PS   Disclosure; Page 122-123; 127pp; English.
XX
CC   The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC   peptides derived from them. The invention is useful for producing an
CC   immune response. It is useful for preventing, treating or ameliorating a
CC   disorder or disease associated with infection of an animal with
CC   Helicobacter. Pharmaceutical composition and vaccines comprising the
CC   sequences of the invention is useful for treating type B gastritis,
CC   peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC   cell lymphoma. The invention is useful as reagents for clinical or
CC   medical diagnosis of Helicobacter infections and for scientific research
CC   on the properties of pathogenicity, virulence and infectivity of
CC   Helicobacter, as well as host defense mechanisms. The present sequence is
CC   Helicobacter sp. PQE/HP30 protein
XX
SQ   Sequence 265 AA;

  Query Match           100.0%;  Score 1279;  DB 5;  Length 265;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-118;
  Matches  253;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 72

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 132

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 192

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 252

Qy        241 ANGDKKSLQIESI 253
              |||||||||||||
Db        253 ANGDKKSLQIESI 265


RESULT 3
AAW20486
ID   AAW20486 standard; protein; 253 AA.
XX
AC   AAW20486;
XX
DT   29-JUL-1997  (first entry)
XX
DE   H. pylori cytoplasmic protein, 4095342.aa.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS   Helicobacter pylori.
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US009122.
XX
PR   07-JUN-1995;   95US-00487032.
PR   01-APR-1996;   96US-00630405.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Smith D,  Berglindh OT,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   N-PSDB; AAT67811.
XX
PT   Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT   useful for vaccines to treat or prevent H. pylori infection, and to
PT   detect Helicobacter.
XX
```

PS Claim 61; Page 651; 1481pp; English.
XX
CC The present sequence is a H. pylori cytoplasmic protein. The protein may
CC be used in a vaccine to prevent or treat H. pylori infection or to
CC identify H. pylori polypeptide binding compounds, useful as potential H.
CC pylori life cycle activators or inhibitors. The genomic sequence of H.
CC pylori (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The sequences were analysed for
CC ORF of at least 180 nucleotides, and the predicted coding regions defined
CC by computer evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF were
CC analysed for significant homology to other known or exported membrane
CC proteins. Having identified and determined the sequences of interest,
CC particular regions can be isolated from H. pylori by PCR amplification
CC for recombinant polypeptide production, e.g. in E. coli hosts
XX
SQ Sequence 253 AA;

Query Match 99.3%; Score 1270; DB 2; Length 253;
Best Local Similarity 99.2%; Pred. No. 1.4e-117;
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              || |||||||||:
Db        241 ANEDKKSLQIESV 253
```

RESULT 4
AAW24673
ID AAW24673 standard; protein; 253 AA.
XX
AC AAW24673;
XX
DT 12-AUG-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 4095342.aa.
XX
KW Transmembrane; cytoplasmic; cell envelope; flagella; transport; secreted;
KW periplasmic; chronic gastritis; duodenal ulcer disease; activator;
KW inhibitor; bacterial life cycle; vaccine; immunise; detection; antisense;
KW inhibition.
XX
OS Helicobacter pylori.
XX
PN WO9719098-A1.
XX
PD 29-MAY-1997.
XX
PF 15-NOV-1996; 96WO-US018542.
XX
PR 17-NOV-1995; 95US-00561469.
XX
PA (ASTR) ASTRA AB.
XX
PI Smith DH;
XX
DR WPI; 1997-298052/27.
DR N-PSDB; AAT77491.
XX
PT Helicobacter pylori nucleic acid sequences and related proteins - used
PT for diagnostics and therapeutics.
XX
PS Claim 18; Page 184; 235pp; English.
XX
CC This sequence represents an H. pylori cytoplasmic protein. Helicobacter
CC pylori has been strongly linked to chronic gastritis and duodenal ulcer
CC disease. The nucleic acid sequences of the invention are used to evaluate
CC compounds, especially activators or inhibitors of bacterial life cycle,
CC for the ability to bind an H. pylori nucleic acid sequence. The nucleic
CC acid sequences, and corresponding proteins, are also useful for
CC generating vaccines for immunising subjects against H. pylori or for use
CC in detecting the presence of Helicobacter species in a sample. Antisense
CC nucleic acid sequences of these sequences are used to inhibit expression
CC of a gene from Helicobacter species. H. pylori whole genomic DNA was
CC isolated and nebulised to a median size of 2000 bp. Purified DNA
CC fragments were blunt-ended and ligated to unique BstXI-linker adapters in
CC 100-1000 fold molar excess. These linkers are complementary to the BstXI-
CC cut pMPX vectors, while the overhang is not self-complementary. Therefore
CC the linkers will not concatemerise nor will the cut vector re-ligate
CC itself easily. The linker-adapter inserts were ligated to each of the 20
CC pMPX vectors to construct a series of shotgun subclone libraries. The
CC purified DNA samples were then sequenced. Note: The ORF/protein reference
CC number for this sequence was obtained from the related specification,
CC WO9640893
XX
SQ Sequence 253 AA;

Query Match 99.3%; Score 1270; DB 2; Length 253;
Best Local Similarity 99.2%; Pred. No. 1.4e-117;
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              || |||||||||:
Db        241 ANEDKKSLQIESV 253
```

RESULT 5
AAW20866
ID AAW20866 standard; protein; 256 AA.
XX
AC AAW20866;
XX
DT 18-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 12ge20305orf30.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis.
XX
OS Helicobacter pylori.
XX
PN WO9640893-A1.

XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US009122.
XX
PR 07-JUN-1995; 95US-00487032.
PR 01-APR-1996; 96US-00630405.
XX
PA (ASTR) ASTRA AB.
XX
PI Smith D, Berglindh OT, Mellgaerd BL;
XX
DR WPI; 1997-052306/05.
DR N-PSDB; AAT68119.
XX
PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT useful for vaccines to treat or prevent H. pylori infection, and to
PT detect Helicobacter.
XX
PS Claim 61; Page 1269; 1481pp; English.
XX
CC This sequence represents a H. pylori cytoplasmic protein. The protein may
CC be used in a vaccine to prevent or treat H. pylori infection or to
CC identify H. pylori polypeptide binding compounds, useful as potential H.
CC pylori life cycle activators or inhibitors. The genomic sequence of H.
CC pylori (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The sequences were analysed for
CC ORF of at least 180 nucleotides, and the predicted coding regions defined
CC by computer evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF were
CC analysed for significant homology to other known or exported membrane
CC proteins. Having identified and determined the sequences of interest,
CC particular regions can be isolated from H. pylori by PCR amplification
CC for recombinant polypeptide production, e.g. in E. coli hosts
XX
SQ Sequence 256 AA;

Query Match 99.3%; Score 1270; DB 2; Length 256;
Best Local Similarity 99.2%; Pred. No. 1.5e-117;
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 63

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 123

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 183

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 243

Qy        241 ANGDKKSLQIESI 253
              || |||||||||:
Db        244 ANEDKKSLQIESV 256
```

RESULT 6
AAE26880
ID AAE26880 standard; protein; 253 AA.
XX
AC AAE26880;
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. pQE/HP30 plasmid DNA encoded protein.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44537.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Disclosure; Page 126-127; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. pQE/HP30 plasmid DNA encoded protein
XX
SQ Sequence 253 AA;

Query Match 98.8%; Score 1264; DB 5; Length 253;
Best Local Similarity 98.8%; Pred. No. 5.6e-117;
Matches 250; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSLEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGDQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              || ||||||||||
Db        241 ANEDKKSLQIESI 253
```

RESULT 7
AAE26879
ID AAE26879 standard; protein; 265 AA.
XX
AC AAE26879;
XX
DT 13-DEC-2002 (first entry)

XX
DE Helicobacter sp. pQE/HP56 plasmid DNA encoded protein.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44536.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Disclosure; Page 124-125; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. pQE/HP56 plasmid DNA encoded protein
XX
SQ Sequence 265 AA;

Query Match 98.8%; Score 1264; DB 5; Length 265;
Best Local Similarity 98.8%; Pred. No. 6e-117;
Matches 250; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db         13 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSLEYKRHGDDYAKYAER 72

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 132

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 192

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GRGLSLAGDQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 252

Qy        241 ANGDKKSLQIESI 253
              || |||||||||||
Db        253 ANEDKKSLQIESI 265
```

RESULT 8
AAW98774
ID AAW98774 standard; protein; 155 AA.
XX
AC AAW98774;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 1170 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US006371.
XX
PR 01-APR-1997; 97US-00833457.
PR 24-JUN-1997; 97US-00881227.
PR 29-JUL-1997; 97US-00902615.
XX
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kleanthous H, Al-Garawi A, Miller C, Tomb J, Oomen RP;
XX
DR WPI; 1998-542293/46.
DR N-PSDB; AAX14493.
XX
PT New isolated Helicobacter polynucleotides - used to develop products for
PT the diagnosis, prevention and treatment of Helicobacter infections and
PT gastrointestinal diseases.
XX
PS Claim 8; Page 1676-1677; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis
XX
SQ Sequence 155 AA;

Query Match 36.9%; Score 472.5; DB 2; Length 155;
Best Local Similarity 63.3%; Pred. No. 1.3e-38;
Matches 93; Conservative 22; Mismatches 29; Indels 3; Gaps 1;

```
Qy         38 NEKLTSSIEYKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKV 97
              || ||:| ||||:| ||||| |||||||:|| ||||:| : ||||||||||| || |||
Db          2 NEDLTWSTEYKRYGHDYAKYPRRIAEELQHYGGNSFANFFRDEGVLYKEILCDACDHLKV 61

Qy         98 NYNKKTETTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNL---NRQALSAATL 154
              |||::: |:|||||||||:|: |||:|   |:||:|:|| : | | :   |:| | |:||
Db         62 NYNEESATSLIEQNMLSKLLKDSLEKMSRREIKELCNELGMTNIDKVIGENKQVLIASTL 121

Qy        155 TLFKMGGFKSYQLAVIVANAVAKTILG 181
              |||| ||  || ||| ||:|: :  ||
Db        122 TLFKAGGSHSYALAVSVADAMVRQTLG 148
```

RESULT 9
AAE26876
ID AAE26876 standard; peptide; 49 AA.
XX
AC AAE26876;
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30-derived peptide #5.
XX

KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44529.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Claim 7; Page 116; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
SQ Sequence 49 AA;

Query Match 19.7%; Score 252; DB 5; Length 49;
Best Local Similarity 100.0%; Pred. No. 2e-17;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 VGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQANGDKKSLQIE 251
||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 VGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQANGDKKSLQIE 49

RESULT 10
AAE26873
ID AAE26873 standard; peptide; 38 AA.
XX
AC AAE26873;
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30-derived peptide #2.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44526.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Claim 7; Page 115; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
SQ Sequence 38 AA;

Query Match 15.5%; Score 198; DB 5; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.2e-12;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 DYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCD 90
||||||||||||||||||||||||||||||||||||||
Db 1 DYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCD 38

RESULT 11
AAE26875
ID AAE26875 standard; peptide; 41 AA.
XX
AC AAE26875;
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30-derived peptide #4.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44528.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.

XX
PS Claim 7; Page 116; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
SQ Sequence 41 AA;

Query Match 15.2%; Score 194; DB 5; Length 41;
Best Local Similarity 100.0%; Pred. No. 8.9e-12;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 NRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGLS 185
||||||||||||||||||||||||||||||||||||||||||||
Db 1 NRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTILGRGLS 41

RESULT 12
AAE26874
ID AAE26874 standard; peptide; 30 AA.
XX
AC AAE26874;
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30-derived peptide #3.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44527.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Claim 7; Page 115; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
SQ Sequence 30 AA;

Query Match 11.9%; Score 152; DB 5; Length 30;
Best Local Similarity 100.0%; Pred. No. 8.5e-08;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 LEEMDDEEVKEMCDELSIKNTDNLNRQALS 150
||||||||||||||||||||||||||||||
Db 1 LEEMDDEEVKEMCDELSIKNTDNLNRQALS 30

RESULT 13
AAE26872
ID AAE26872 standard; peptide; 30 AA.
XX
AC AAE26872;
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30-derived peptide #1.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44525.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Claim 7; Page 115; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
SQ Sequence 30 AA;

Query Match 11.7%; Score 150; DB 5; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFG 30
||||||||||||||||||||||||||||||

```
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFG 30
```

RESULT 14
AAB98720
ID AAB98720 standard; protein; 222 AA.
XX
AC AAB98720;
XX
DT 29-AUG-2001 (first entry)
XX
DE Human breast cancer-associated protein BC-2, SEQ ID NO:16.
XX
KW Human breast cancer-associated protein; BC-2; marker protein;
KW nuclear matrix protein; diagnosis; detection; tumour.
XX
OS Homo sapiens.
XX
PN US6218131-B1.
XX
PD 17-APR-2001.
XX
PF 06-OCT-1997; 97US-00944604.
XX
PR 05-JUN-1996; 96US-00658639.
XX
PA (MATR-) MATRITECH INC.
XX
PI Keesee SK, Obar R, Wu Y;
XX
DR WPI; 2001-396355/42.
DR N-PSDB; AAH25902.
XX
PT Diagnosing or detecting breast cancer in an individual comprises
PT detecting the presence of breast cancer-associated proteins in a
PT biological sample.
XX
PS Claim 1; Col 27-30; 30pp; English.
XX
CC The invention relates to novel human breast cancer-associated proteins,
CC and their use in diagnosing and detecting breast cancer. The breast
CC cancer-associated proteins of the invention are nuclear matrix proteins
CC designated BC-2 (AAB98720), BC-8 isoform A (BC-8A, AAB98721) and CC BC-8
CC isoform B (BC-8B, AAB98722), although 6 other breast cancer-associated
CC nuclear matrix proteins (BC-1, BC-3, BC-4, BC-5, BC-6, and BC-7) were
CC also isolated (sequences not given in the specification). The novel
CC breast cancer-associated proteins are present in the nuclear matrix of
CC breast cancer cells, but are not present in the nuclear matrix of cells
CC from normal breast tissue. The invention also encompasses fragments of
CC the breast cancer-associated proteins (AAB98709-AAB98716), and methods
CC for their use in breast cancer diagnosis. The breast cancer marker
CC proteins of the invention or fragments thereof may be detected in a
CC sample of breast tissue or blood from an individual. Isolated breast
CC cancer-associated proteins may also be used to prepare antibodies,
CC particularly monoclonal antibodies, which may be used to detect or
CC diagnose breast cancer. Nucleic acids encoding the breast cancer-
CC associated proteins may also be used in the diagnosis and detection of
CC breast cancer, and in the isolation of DNA or protein sequences which may
CC interact with breast cancer-associated nuclear matrix proteins. The
CC present sequence represents the full-length breast cancer-associated
CC protein BC-2
XX
SQ Sequence 222 AA;

Query Match 7.7%; Score 99; DB 4; Length 222;
Best Local Similarity 21.4%; Pred. No. 0.27;
Matches 39; Conservative 48; Mismatches 83; Indels 12; Gaps 5;

```
Qy          6 DRDLEFLKQLESSDLLDLFEVLVFGK-DGEKRHNEKLTSSIEYKRHGDDYAKYAERIAEE 64
              ||: : |:  |   : |: ::   |: |  :   : | :  | |         : :: :
Db         30 DRERQKLETQEKKIIADIKKMAKQGQMDAVRIMAKDLVRTRRYVRKFVLMRANIQAVSLK 89

Qy         65 LQYYGS-NSFASFIKG----EGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILER 119
```

```
Db         90 IQTLKSNNSMAQAMKGVTKAMGTMNRQLKLPQIQKIMMEFERQAEIMDMKEEMMNDAIDD 149

Qy        120 SLEEMDDEE-----VKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANA 174
              :: : :|||     | :: ||| :  || |:    :  :|::   || |:   |  :|:|
Db        150 AMGDEEDEEESDAVVSQVLDELGLSLTDELSNLPSTGGSLSV-AAGGKKAEAAASALADA 208

Qy        175 VA 176
              |
Db        209 DA 210
```

RESULT 15
ABP40434
ID ABP40434 standard; protein; 274 AA.
XX
AC ABP40434;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5279.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-00134001.
XX
PR 14-AUG-1997; 97US-0055779P.
PR 08-NOV-1997; 97US-0064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2002-381255/41.
DR N-PSDB; ABN92979.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT polypeptide, useful for diagnosing and treating bacterial infections.
XX
PS Disclosure; SEQ ID NO 5279; 267pp; English.
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
CC also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life cycle
CC or inhibit S. epidermidis infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site
XX
SQ Sequence 274 AA;

Query Match 7.7%; Score 99; DB 5; Length 274;
Best Local Similarity 23.8%; Pred. No. 0.37;
Matches 62; Conservative 44; Mismatches 88; Indels 66; Gaps 14;

```
Qy          3 YKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSS----------------IE 46
              | |: ||                |:  | |:: : : : :|                :|
Db          4 YNYNNDL----------------FLILKKFGDEMNKDTIETSREALIKDVVMIAARILLE 47

Qy         47 YKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKT-ET 105
                    |         ||| :| |  |||| :    | ||: |    :        |  || :|
Db         48 SGAEGTRVEDTMARIATKLGYPESNSFVTNTVIEFVLHNEAYPRL-------YRIKTRDT 100
```

```
Qy        106 TLIEQNMLSKILERSLE-EMDDEEVKEMCDELSI-KNTDNLNRQALSA---ATLTLFKMG 160
              ||: :  ::|  :     |  || |   :|: : |   :|  : ::|   || |:  |
Db        101 NLIKISQANEISRQITNGTMTLEEAKYQLEEIYVAKRDSSLPFKGIAAAIIATSFLYLQG 160

Qy        161 GFKSYQLAVIVANAVAKTILGRGLSLAGNQVLTRTL--SFLTGPVGWIITGVWTAIDIAG 218
              |   :|  |:   : ||   |  :   ::| | |   |:   :| :: |:   | : |
Db        161 G----RLVDIITAWLAGTI---GYLVV--EILDRKLHAQFIPEFIGSLVIGI---ISVIG 208

Qy        219 PAYRVTIP----ACIVVATL 234
              :   :|    | |::| :
Db        209 HAF---VPSGDLATIIIAAV 225
```

Search completed: July 5, 2004, 03:11:25
Job time : 40 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 5, 2004, 03:10:43 ; Search time 14 Seconds
(without alignments)
1738.318 Million cell updates/sec

Title: US-09-732-091-4
Perfect score: 1279
Sequence: 1 MAYKYDRDLEFLKQLESSDL.....LRLKTQQANGDKKSLQIESI 253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1279	100.0	253	2	D64718	conserved hypothet
2	1270	99.3	253	2	B71800	hypothetical prote
3	784.5	61.3	207	2	E64718	conserved hypothet
4	630	49.3	209	2	A71800	hypothetical prote
5	472.5	36.9	155	2	C64718	conserved hypothet
6	265	20.7	237	2	C64721	hypothetical prote
7	265	20.7	237	2	AC0503	conserved hypothet
8	262	20.5	237	2	C90630	probable oxidoredu
9	262	20.5	237	2	C85481	probable oxidoredu
10	133	10.4	39	2	F64718	hypothetical prote
11	98.5	7.7	1009	2	C89910	hypothetical prote
12	98	7.7	282	2	AE0360	hypothetical prote
13	98	7.7	1099	2	G90546	conserved hypothet
14	97.5	7.6	311	2	C81380	probable D-2-hydro
15	97.5	7.6	728	2	A81385	probable ATP /GTP
16	97	7.6	916	2	E71330	probable preprotei
17	96.5	7.5	373	2	G70355	8-amino-7-oxononan
18	96	7.5	262	2	T33071	hypothetical prote
19	96	7.5	420	2	F96920	probable permease
20	95.5	7.5	344	2	A47025	abortive phage res
21	95	7.4	517	2	C81327	hypothetical prote
22	95	7.4	2649	2	A40937	bullous pemphigoid
23	93.5	7.3	802	2	S49252	penicillin amidase
24	93	7.3	355	2	E90488	transposase ISC122
25	93	7.3	610	2	G70178	exodeoxyribonuclea
26	92.5	7.2	430	2	T33070	hypothetical prote
27	92.5	7.2	802	2	I39665	penicillin amidase
28	92.5	7.2	886	2	T16536	hypothetical prote
29	92	7.2	308	2	H81345	hypothetical prote
30	92	7.2	1939	2	T18372	repeat organellar
31	91.5	7.2	399	2	G72253	RNA polymerase sig
32	91.5	7.2	472	2	A64320	PET112 homolog - M
33	91.5	7.2	508	2	T24622	hypothetical prote
34	91.5	7.2	615	2	S42797	rad 26 protein - f
35	91.5	7.2	1769	2	S53378	probable membrane
36	91.5	7.2	1847	2	E64477	replication factor
37	91	7.1	284	2	A45488	body-wall muscle t
38	91	7.1	359	2	AI2368	hypothetical prote
39	90.5	7.1	320	2	B97214	hypothetical prote
40	90.5	7.1	1017	2	D90550	vsaa-like (mycopla
41	90.5	7.1	1411	2	T18417	hypothetical prote
42	90.5	7.1	1417	2	T18418	hypothetical prote
43	90	7.0	404	1	S03849	ribonucleoprotein
44	89.5	7.0	280	2	A72046	conserved hypothet
45	89.5	7.0	280	2	F86578	CT671 hypothetical

ALIGNMENTS

RESULT 1
D64718
conserved hypothetical protein HP1588 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: D64718
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKer
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: D64718
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-253 <TOM>
A;Cross-references: GB:AE000656; GB:AE000511; NID:g2314771; PIDN:AAD08627.1; PID:g2314

Query Match 100.0%; Score 1279; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 4e-88;
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSPLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSPLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              |||||||||||||
Db        241 ANGDKKSLQIESI 253
```

RESULT 2
B71800
hypothetical protein jhp1494 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C;Accession: B71800
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: B71800
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <ARN>
A;Cross-references: GB:AE001571; GB:AE001439; NID:g4156120; PIDN:AAD07073.1; PID:g415612
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp1494

Query Match 99.3%; Score 1270; DB 2; Length 253;
Best Local Similarity 99.2%; Pred. No. 1.9e-87;
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAYKYDRDLEFLKQLESSDLLDLFEVLVFGKDGEKRHNEKLTSSIEYKRHGDDYAKYAER 60

Qy         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYNKKTETTLIEQNMLSKILERS 120

Qy        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMGGFKSYQLAVIVANAVAKTIL 180

Qy        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPAYRVTIPACIVVATLRLKTQQ 240

Qy        241 ANGDKKSLQIESI 253
              || |||||||||:
Db        241 ANEDKKSLQIESV 253
```

RESULT 3
E64718
conserved hypothetical protein HP1589 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Aug-1997
C;Accession: E64718
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: E64718
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-207 <TOM>
A;Cross-references: GB:AE000511; TIGR:HP1589
C;Genetics:
A;Start codon: TTG

Query Match 61.3%; Score 784.5; DB 2; Length 207;
Best Local Similarity 73.6%; Pred. No. 2.2e-51;
Matches 153; Conservative 21; Mismatches 33; Indels 1; Gaps 1;

```
Qy         41 LTSSIEYKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKVNYN 100
              :||| ||:|:  |||||  ||||||| || ||| :| : ||||||||||| || ||||||
Db          1 MTSSTEYQRYGYDYAKYPRRIAEELQRYGGNSFMNFFRDEGVLYKEILCDACDHLKVNYN 60

Qy        101 KKTETTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNLNRQALSAATLTLFKMG 160
              ||::|||||:|||| ||::|||:| |||::|:|||| :|||: | :|||| | ||||:||
Db         61 KKSDTTLIEENMLSSILQKSLEKMSDEEIRELCDELGVKNTNKLGKQALSTAALTLFRMG 120

Qy        161 GFKSYQLAVIVANAVAKTILGRGLSLAGNQVLTRTLSFLTGPVGWIITGVWTAIDIAGPA 220
              ||||||||:|||||| | |  |||||  |  ||| || ||||:|||||||||||||||||
```

```
Db        121 GFKSYQLALIVANAVIKAIFQRGLSLGANAALTRGLSILTGPIGWIITGVWTAIDIAGPA 180

Qy        221 YRVTIPACIVVATLRLKTQQANGDKKSL 248
              |||||||||:|||||||  |||  |  |
Db        181 YRVTIPACILVATLRLKA-QANEIKNIL 207
```

RESULT 4
A71800
hypothetical protein jhp1493 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C;Accession: A71800
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: A71800
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-209 <ARN>
A;Cross-references: GB:AE001571; GB:AE001439; NID:g4156120; PIDN:AAD07072.1; PID:g415
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp1493

Query Match 49.3%; Score 630; DB 2; Length 209;
Best Local Similarity 59.8%; Pred. No. 7.4e-40;
Matches 122; Conservative 32; Mismatches 46; Indels 4; Gaps 2;

```
Qy         38 NEKLTSSIEYKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKV 97
              ||:|||  ||:|:| |||||  ||||||| || ||||:| : |||||||||| || | :
Db          2 NEELTSLTEYQRYGHDYAKYPRRIAEELQRYGGNSFANFFRDEGVLYKEILCDACDHLDI 61

Qy         98 NYNKKTETTLIEQNMLSKILERSLEEMDDEEVKEMCDELSIKNTDNL---NRQALSAATL 154
              |||::: |:|||||||||:|: ||||:|   |:||:|| | : | | :   |:| | |: |
Db         62 NYNERSATSLIEQNMLSKLLKDSLEKMSGREIKELCDGLGMPNIDKVIGENKQVLIASVL 121

Qy        155 TLFKMGGFKSYQLAVIVANAVAKTILGRGL-SLAGNQVLTRTLSFLTGPVGWIITGVWTA 213
              |||| ||  || ||| ||:|: :  || || |: |   | :||  | ||:||:|||    :
Db        122 TLFKAGGSHSYALAVAVADAMVRQTLGHGLSSVVGKVALKKTLDILAGPIGWVITGALVS 181

Qy        214 IDIAGPAYRVTIPACIVVATLRLK 237
              |::|||||||:|||::|||| |
Db        182 INLAGPAYRVTVPACVLVATLRKK 205
```

RESULT 5
C64718
conserved hypothetical protein HP1587 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: C64718
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: C64718
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-155 <TOM>
A;Cross-references: GB:AE000656; GB:AE000511; NID:g2314771; PIDN:AAD08626.1; PID:g231

Query Match 36.9%; Score 472.5; DB 2; Length 155;
Best Local Similarity 63.3%; Pred. No. 2.8e-28;
Matches 93; Conservative 22; Mismatches 29; Indels 3; Gaps 1;

```
Qy         38 NEKLTSSIEYKRHGDDYAKYAERIAEELQYYGSNSFASFIKGEGVLYKEILCDVCDKLKV 97
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 5, 2004, 01:15:47 ; Search time 64 Seconds
(without alignments)
6701.674 Million cell updates/sec

Title: US-09-732-091-3
Perfect score: 1349
Sequence: 1 atggcatacaaatatgatag.....cgttgcaaatagaatccatt 759

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09732091/runat_02072004_182406_18996/app_query.fasta_1.903
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09732091_@CGN_1_1_101_@runat_02072004_182406_18996 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=2 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1279	94.8	253	5	AAE26860	Aae26860 Helicobac
2	1279	94.8	265	5	AAE26878	Aae26878 Helicobac
3	1270	94.1	253	2	AAW20486	Aaw20486 H. pylori
4	1270	94.1	253	2	AAW24673	Aaw24673 H. pylori
5	1270	94.1	256	2	AAW20866	Aaw20866 H. pylori
6	1264	93.7	253	5	AAE26880	Aae26880 Helicobac
7	1264	93.7	265	5	AAE26879	Aae26879 Helicobac
8	472.5	35.0	155	2	AAW98774	Aaw98774 H. pylori
9	252	18.7	49	5	AAE26876	Aae26876 Helicobac
10	198	14.7	38	5	AAE26873	Aae26873 Helicobac
11	194	14.4	41	5	AAE26875	Aae26875 Helicobac
12	152	11.3	30	5	AAE26874	Aae26874 Helicobac
13	150	11.1	30	5	AAE26872	Aae26872 Helicobac
14	101.5	7.5	173	2	AAW44365	Aaw44365 Breast ca
15	101.5	7.5	173	4	AAB98717	Aab98717 Human bre
16	99	7.3	222	4	AAB98720	Aab98720 Human bre
17	99	7.3	274	5	ABP40434	Abp40434 Staphyloc
18	98.5	7.3	1009	6	ABU15908	Abu15908 Protein e
19	98	7.3	413	6	ABM71523	Abm71523 Staphyloc
20	97	7.2	916	6	ABU48566	Abu48566 Protein e
c 21	96.5	7.2	414	7	ADC07898	Adc07898 Rice prot
22	96.5	7.2	701	5	ABP73563	Abp73563 Candida a
23	95	7.0	517	6	ABU26619	Abu26619 Protein e
24	95	7.0	2125	5	ABG77172	Abg77172 Prostate
25	95	7.0	2649	7	ADB79929	Adb79929 Human put
26	93.5	6.9	1196	6	ABU24813	Abu24813 Protein e
27	93	6.9	381	4	AAG82981	Aag82981 S. epider
28	93	6.9	610	6	ABU19293	Abu19293 Protein e
c 29	93	6.9	667	2	AAW94919	Aaw94919 Rat phero
30	92.5	6.9	996	4	AAU33755	Aau33755 Staphyloc
31	92.5	6.9	1009	4	AAU36548	Aau36548 Staphyloc
32	92	6.8	309	5	ABP30227	Abp30227 Streptoco
33	92	6.8	327	5	ABP28162	Abp28162 Streptoco
34	91.5	6.8	187	3	AAG44550	Aag44550 Arabidops
35	91.5	6.8	187	3	AAG48726	Aag48726 Arabidops
36	91.5	6.8	241	3	AAG44549	Aag44549 Arabidops
37	91.5	6.8	241	3	AAG48725	Aag48725 Arabidops
38	91.5	6.8	256	3	AAG48724	Aag48724 Arabidops
39	91.5	6.8	256	3	AAG44548	Aag44548 Arabidops
40	91.5	6.8	375	5	ABP30479	Abp30479 Streptoco
41	91.5	6.8	390	5	ABP27884	Abp27884 Streptoco
42	91.5	6.8	1769	6	ABR52974	Abr52974 Protein s
43	91.5	6.8	1847	3	AAY51631	Aay51631 M. jannas
44	91.5	6.8	1847	3	AAY52002	Aay52002 M. jannas
45	90.5	6.7	1411	2	AAY31948	Aay31948 Plasmodiu

ALIGNMENTS

RESULT 1
AAE26860
ID AAE26860 standard; protein; 253 AA.
XX
AC AAE26860;
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30 protein.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44513.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.

XX
PS Claim 1; Page 111-112; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30 protein
XX
SQ Sequence 253 AA;

Alignment Scores:

Pred. No.:	6.71e-135	Length:	253
Score:	1279.00	Matches:	253
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	94.81%	Indels:	0
DB:	5	Gaps:	0

US-09-732-091-3 (1-759) x AAE26860 (1-253)

```
Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              |||||||||||||||||||||||||||||||||||||||
Db        241 AlaAsnGlyAspLysLysSerLeuGlnIleGluSerIle 253
```

RESULT 2
AAE26878
ID AAE26878 standard; protein; 265 AA.
XX
AC AAE26878;
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. PQE/HP30 protein.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
FH Key Location/Qualifiers
FT Misc-difference 12
FT /note= "Encoded by TTC"
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44535.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Disclosure; Page 122-123; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. PQE/HP30 protein
XX
SQ Sequence 265 AA;

Alignment Scores:

Pred. No.:	6.84e-135	Length:	265
Score:	1279.00	Matches:	253
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	94.81%	Indels:	0
DB:	5	Gaps:	0

US-09-732-091-3 (1-759) x AAE26878 (1-265)

```
Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 32

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 52

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         53 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 72

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 92

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 112

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        113 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 132

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 152

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 172

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 192

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 212

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 232

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 252

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              |||||||||||||||||||||||||||||||||||||||
Db        253 AlaAsnGlyAspLysLysSerLeuGlnIleGluSerIle 265
```

RESULT 3
AAW20486
ID AAW20486 standard; protein; 253 AA.
XX
AC AAW20486;
XX
DT 29-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 4095342.aa.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US009122.
XX
PR 07-JUN-1995; 95US-00487032.
PR 01-APR-1996; 96US-00630405.
XX
PA (ASTR) ASTRA AB.
XX
PI Smith D, Berglindh OT, Mellgaerd BL;
XX
DR WPI; 1997-052306/05.
DR N-PSDB; AAT67811.
XX
PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT useful for vaccines to treat or prevent H. pylori infection, and to
PT detect Helicobacter.
XX
PS Claim 61; Page 651; 1481pp; English.
XX
CC The present sequence is a H. pylori cytoplasmic protein. The protein may
CC be used in a vaccine to prevent or treat H. pylori infection or to
CC identify H. pylori polypeptide binding compounds, useful as potential H.
CC pylori life cycle activators or inhibitors. The genomic sequence of H.
CC pylori (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The sequences were analysed for
CC ORF of at least 180 nucleotides, and the predicted coding regions defined
CC by computer evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF were
CC analysed for significant homology to other known or exported membrane
CC proteins. Having identified and determined the sequences of interest,
CC particular regions can be isolated from H. pylori by PCR amplification
CC for recombinant polypeptide production, e.g. in E. coli hosts
XX
SQ Sequence 253 AA;

Alignment Scores:

Pred. No.:	6.94e-134	Length:	253
Score:	1270.00	Matches:	251
Percent Similarity:	99.60%	Conservative:	1
Best Local Similarity:	99.21%	Mismatches:	1
Query Match:	94.14%	Indels:	0
DB:	2	Gaps:	0

US-09-732-091-3 (1-759) x AAW20486 (1-253)

```
Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              ||||||   |||||||||||||||||||||||||:::
Db        241 AlaAsnGluAspLysLysSerLeuGlnIleGluSerVal 253

RESULT 4
AAW24673
ID   AAW24673 standard; protein; 253 AA.
XX
AC   AAW24673;
XX
DT   12-AUG-1997  (first entry)
XX
DE   H. pylori cytoplasmic protein, 4095342.aa.
XX
KW   Transmembrane; cytoplasmic; cell envelope; flagella; transport; secreted;
KW   periplasmic; chronic gastritis; duodenal ulcer disease; activator;
KW   inhibitor; bacterial life cycle; vaccine; immunise; detection; antisense;
KW   inhibition.
XX
OS   Helicobacter pylori.
XX
PN   WO9719098-A1.
XX
PD   29-MAY-1997.
XX
PF   15-NOV-1996;   96WO-US018542.
XX
PR   17-NOV-1995;   95US-00561469.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Smith DH;
XX
DR   WPI; 1997-298052/27.
DR   N-PSDB; AAT77491.
XX
PT   Helicobacter pylori nucleic acid sequences and related proteins - used
PT   for diagnostics and therapeutics.
XX
PS   Claim 18; Page 184; 235pp; English.
XX
CC   This sequence represents an H. pylori cytoplasmic protein. Helicobacter
CC   pylori has been strongly linked to chronic gastritis and duodenal ulcer
CC   disease. The nucleic acid sequences of the invention are used to evaluate
CC   compounds, especially activators or inhibitors of bacterial life cycle,
CC   for the ability to bind an H. pylori nucleic acid sequence. The nucleic
CC   acid sequences, and corresponding proteins, are also useful for
CC   generating vaccines for immunising subjects against H. pylori or for use
CC   in detecting the presence of Helicobacter species in a sample. Antisense
CC   nucleic acid sequences of these sequences are used to inhibit expression
CC   of a gene from Helicobacter species. H. pylori whole genomic DNA was
CC   isolated and nebulised to a median size of 2000 bp. Purified DNA
CC   fragments were blunt-ended and ligated to unique BstXI-linker adapters in
CC   100-1000 fold molar excess. These linkers are complementary to the BstXI-
CC   cut pMPX vectors, while the overhang is not self-complementary. Therefore
CC   the linkers will not concatemerise nor will the cut vector re-ligate
CC   itself easily. The linker-adapter inserts were ligated to each of the 20
CC   pMPX vectors to construct a series of shotgun subclone libraries. The
CC   purified DNA samples were then sequenced. Note: The ORF/protein reference
CC   number for this sequence was obtained from the related specification,
CC   WO9640893
XX
SQ   Sequence 253 AA;

Alignment Scores:
Pred. No.:                6.94e-134      Length:          253
Score:                    1270.00        Matches:         251
Percent Similarity:       99.60%         Conservative:    1
Best Local Similarity:    99.21%         Mismatches:      1
Query Match:              94.14%         Indels:          0
DB:                       2              Gaps:            0

US-09-732-091-3 (1-759) x AAW24673 (1-253)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              ||||||   |||||||||||||||||||||||||:::
Db        241 AlaAsnGluAspLysLysSerLeuGlnIleGluSerVal 253

RESULT 5
AAW20866
```

```
ID   AAW20866 standard; protein; 256 AA.
XX
AC   AAW20866;
XX
DT   18-JUL-1997  (first entry)
XX
DE   H. pylori cytoplasmic protein, 12ge20305orf30.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis.
XX
OS   Helicobacter pylori.
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US009122.
XX
PR   07-JUN-1995;   95US-00487032.
PR   01-APR-1996;   96US-00630405.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Smith D,  Berglindh OT,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   N-PSDB; AAT68119.
XX
PT   Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT   useful for vaccines to treat or prevent H. pylori infection, and to
PT   detect Helicobacter.
XX
PS   Claim 61; Page 1269; 1481pp; English.
XX
CC   This sequence represents a H. pylori cytoplasmic protein. The protein may
CC   be used in a vaccine to prevent or treat H. pylori infection or to
CC   identify H. pylori polypeptide binding compounds, useful as potential H.
CC   pylori life cycle activators or inhibitors. The genomic sequence of H.
CC   pylori (ATCC 55679) was determined from overlapping contigs generated by
CC   mechanically shearing the bacterial DNA. The sequences were analysed for
CC   ORF of at least 180 nucleotides, and the predicted coding regions defined
CC   by computer evaluation. To identify likely H. pylori antigens for vaccine
CC   development, the amino acid sequences predicted from various ORF were
CC   analysed for significant homology to other known or exported membrane
CC   proteins. Having identified and determined the sequences of interest,
CC   particular regions can be isolated from H. pylori by PCR amplification
CC   for recombinant polypeptide production, e.g. in E. coli hosts
XX
SQ   Sequence 256 AA;

Alignment Scores:
Pred. No.:                6.98e-134      Length:          256
Score:                    1270.00        Matches:         251
Percent Similarity:       99.60%         Conservative:    1
Best Local Similarity:    99.21%         Mismatches:      1
Query Match:              94.14%         Indels:          0
DB:                       2              Gaps:            0

US-09-732-091-3 (1-759) x AAW20866 (1-256)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 23

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 43

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         44 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 63

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 83

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 103

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 123

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 143

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 163

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 183

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 203

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 223

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 243

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              ||||||   ||||||||||||||||||||||||:::
Db        244 AlaAsnGluAspLysLysSerLeuGlnIleGluSerVal 256

RESULT 6
AAE26880
ID   AAE26880 standard; protein; 253 AA.
XX
AC   AAE26880;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Helicobacter sp. pQE/HP30 plasmid DNA encoded protein.
XX
KW   HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW   type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW   low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW   immunostimulant; cytostatic; peptic ulcer.
XX
OS   Helicobacter sp.
XX
PN   WO200251237-A2.
XX
PD   04-JUL-2002.
XX
PF   07-DEC-2001; 2001WO-US048392.
XX
PR   07-DEC-2000; 2000US-00732091.
XX
PA   (ANTE-) ANTEX BIOLOGICS INC.
XX
PI   Tian J,  Walker R,  Jackson WJ;
XX
DR   WPI; 2002-666854/71.
DR   N-PSDB; AAD44537.
```

```
XX
PT   Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT   the proteins, useful as vaccines for raising immune response in animals.
XX
PS   Disclosure; Page 126-127; 127pp; English.
XX
CC   The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC   peptides derived from them. The invention is useful for producing an
CC   immune response. It is useful for preventing, treating or ameliorating a
CC   disorder or disease associated with infection of an animal with
CC   Helicobacter. Pharmaceutical composition and vaccines comprising the
CC   sequences of the invention is useful for treating type B gastritis,
CC   peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC   cell lymphoma. The invention is useful as reagents for clinical or
CC   medical diagnosis of Helicobacter infections and for scientific research
CC   on the properties of pathogenicity, virulence and infectivity of
CC   Helicobacter, as well as host defense mechanisms. The present sequence is
CC   Helicobacter sp. pQE/HP30 plasmid DNA encoded protein
XX
SQ   Sequence 253 AA;

Alignment Scores:
Pred. No.:              3.3e-133      Length:          253
Score:                  1264.00       Matches:         250
Percent Similarity:     99.60%        Conservative:    2
Best Local Similarity:  98.81%        Mismatches:      1
Query Match:            93.70%        Indels:          0
DB:                     5             Gaps:            0

US-09-732-091-3 (1-759) x AAE26880 (1-253)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuThrSerSerLeuGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        181 GlyArgGlyLeuSerLeuAlaGlyAspGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              ||||||   ||||||||||||||||||||||||||||||
Db        241 AlaAsnGluAspLysLysSerLeuGlnIleGluSerIle 253

RESULT 7
AAE26879
ID   AAE26879 standard; protein; 265 AA.
XX
AC   AAE26879;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Helicobacter sp. pQE/HP56 plasmid DNA encoded protein.
XX
KW   HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW   type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW   low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW   immunostimulant; cytostatic; peptic ulcer.
XX
OS   Helicobacter sp.
XX
PN   WO200251237-A2.
XX
PD   04-JUL-2002.
XX
PF   07-DEC-2001; 2001WO-US048392.
XX
PR   07-DEC-2000; 2000US-00732091.
XX
PA   (ANTE-) ANTEX BIOLOGICS INC.
XX
PI   Tian J,  Walker R,  Jackson WJ;
XX
DR   WPI; 2002-666854/71.
DR   N-PSDB; AAD44536.
XX
PT   Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT   the proteins, useful as vaccines for raising immune response in animals.
XX
PS   Disclosure; Page 124-125; 127pp; English.
XX
CC   The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC   peptides derived from them. The invention is useful for producing an
CC   immune response. It is useful for preventing, treating or ameliorating a
CC   disorder or disease associated with infection of an animal with
CC   Helicobacter. Pharmaceutical composition and vaccines comprising the
CC   sequences of the invention is useful for treating type B gastritis,
CC   peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC   cell lymphoma. The invention is useful as reagents for clinical or
CC   medical diagnosis of Helicobacter infections and for scientific research
CC   on the properties of pathogenicity, virulence and infectivity of
CC   Helicobacter, as well as host defense mechanisms. The present sequence is
CC   Helicobacter sp. pQE/HP56 plasmid DNA encoded protein
XX
SQ   Sequence 265 AA;

Alignment Scores:
Pred. No.:              3.36e-133     Length:          265
Score:                  1264.00       Matches:         250
Percent Similarity:     99.60%        Conservative:    2
Best Local Similarity:  98.81%        Mismatches:      1
Query Match:            93.70%        Indels:          0
DB:                     5             Gaps:            0

US-09-732-091-3 (1-759) x AAE26879 (1-265)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
```

```
Db          13 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 32

Qy          61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120

Db          33 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 52

Qy         121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180

Db          53 LeuThrSerSerLeuGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 72

Qy         181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240

Db          73 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 92

Qy         241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300

Db          93 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 112

Qy         301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360

Db         113 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 132

Qy         361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420

Db         133 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 152

Qy         421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480

Db         153 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 172

Qy         481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540

Db         173 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 192

Qy         541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600

Db         193 GlyArgGlyLeuSerLeuAlaGlyAspGlnValLeuThrArgThrLeuSerPheLeuThr 212

Qy         601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660

Db         213 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 232

Qy         661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720

Db         233 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 252

Qy         721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759

Db         253 AlaAsnGluAspLysLysSerLeuGlnIleGluSerIle 265

RESULT 8
AAW98774
ID   AAW98774 standard; protein; 155 AA.
XX
AC   AAW98774;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 1170 protein.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
XX
OS   Helicobacter pylori.
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US006371.
XX
PR   01-APR-1997;   97US-00833457.
PR   24-JUN-1997;   97US-00881227.
PR   29-JUL-1997;   97US-00902615.
XX
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Kleanthous H,  Al-Garawi A,  Miller C,  Tomb J,  Oomen RP;
XX
DR   WPI; 1998-542293/46.
DR   N-PSDB; AAX14493.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products for
PT   the diagnosis, prevention and treatment of Helicobacter infections and
PT   gastrointestinal diseases.
XX
PS   Claim 8; Page 1676-1677; 2054pp; English.
XX
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic, and atrophic gastritis, and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis
XX
SQ   Sequence 155 AA;

Alignment Scores:
Pred. No.:                4.64e-44        Length:          155
Score:                    472.50          Matches:         93
Percent Similarity:       78.23%          Conservative:    22
Best Local Similarity:    63.27%          Mismatches:      29
Query Match:              35.03%          Indels:          3
DB:                       2               Gaps:            1

US-09-732-091-3 (1-759) x AAW98774 (1-155)

Qy         112 AATGAAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATAC 171

Db           2 AsnGluAspLeuThrAsnSerThrGluTyrLysArgTyrGlyHisAspTyrAlaLysTyr 21

Qy         172 GCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATT 231

Db          22 ProArgArgIleAlaGluGluLeuGlnHisTyrGlyGlyAsnSerPheAlaAsnPhePhe 41

Qy         232 AAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTC 291

Db          42 ArgAspGluGlyValLeuTyrLysGluIleLeuCysAspAlaCysAspHisLeuLysVal 61

Qy         292 AATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTA 351

Db          62 AsnTyrAsnGluGluSerAlaThrSerLeuIleGluGlnAsnMetLeuSerLysLeuLeu 81

Qy         352 GAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCC 411

Db          82 LysAspSerLeuGluLysMetSerArgArgGluIleLysGluLeuCysAsnGluLeuGly 101

Qy         412 ATAAAAAACACGGACAATTTA---------AACAGACAAGCCTTAAGCGCGGCGACTTTA 462

Db         102 MetThrAsnIleAspLysValIleGlyGluAsnLysGlnValLeuIleAlaSerThrLeu 121

Qy         463 ACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCG 522

Db         122 ThrLeuPheLysAlaGlyGlySerHisSerTyrAlaLeuAlaValSerValAlaAspAla 141

Qy         523 GTCGCAAAAACCATTCTAGGG 543

Db         142 MetValArgGlnThrLeuGly 148

RESULT 9
AAE26876
ID   AAE26876 standard; peptide; 49 AA.
XX
```

```
AC   AAE26876;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Helicobacter sp. HP30-derived peptide #5.
XX
KW   HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW   type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW   low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW   immunostimulant; cytostatic; peptic ulcer.
XX
OS   Helicobacter sp.
XX
PN   WO200251237-A2.
XX
PD   04-JUL-2002.
XX
PF   07-DEC-2001; 2001WO-US048392.
XX
PR   07-DEC-2000; 2000US-00732091.
XX
PA   (ANTE-) ANTEX BIOLOGICS INC.
XX
PI   Tian J,  Walker R,  Jackson WJ;
XX
DR   WPI; 2002-666854/71.
DR   N-PSDB; AAD44529.
XX
PT   Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT   the proteins, useful as vaccines for raising immune response in animals.
XX
PS   Claim 7; Page 116; 127pp; English.
XX
CC   The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC   peptides derived from them. The invention is useful for producing an
CC   immune response. It is useful for preventing, treating or ameliorating a
CC   disorder or disease associated with infection of an animal with
CC   Helicobacter. Pharmaceutical composition and vaccines comprising the
CC   sequences of the invention is useful for treating type B gastritis,
CC   peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC   cell lymphoma. The invention is useful as reagents for clinical or
CC   medical diagnosis of Helicobacter infections and for scientific research
CC   on the properties of pathogenicity, virulence and infectivity of
CC   Helicobacter, as well as host defense mechanisms. The present sequence is
CC   Helicobacter sp. HP30-derived peptide
XX
SQ   Sequence 49 AA;

Alignment Scores:
Pred. No.:                2.12e-19        Length:          49
Score:                    252.00          Matches:         49
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              18.68%          Indels:          0
DB:                       5               Gaps:            0

US-09-732-091-3 (1-759) x AAE26876 (1-49)

Qy        607 GTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAlaTyrArg 20

Qy        667 GTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAAGCCAAT 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGlnAlaAsn 40

Qy        727 GGAGATAAGAAGTCGTTGCAAATAGAA 753
              ||||||||||||||||||||||||||||
Db         41 GlyAspLysLysSerLeuGlnIleGlu 49

RESULT 10
AAE26873
ID   AAE26873 standard; peptide; 38 AA.
XX
AC   AAE26873;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Helicobacter sp. HP30-derived peptide #2.
XX
KW   HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW   type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW   low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW   immunostimulant; cytostatic; peptic ulcer.
XX
OS   Helicobacter sp.
XX
PN   WO200251237-A2.
XX
PD   04-JUL-2002.
XX
PF   07-DEC-2001; 2001WO-US048392.
XX
PR   07-DEC-2000; 2000US-00732091.
XX
PA   (ANTE-) ANTEX BIOLOGICS INC.
XX
PI   Tian J,  Walker R,  Jackson WJ;
XX
DR   WPI; 2002-666854/71.
DR   N-PSDB; AAD44526.
XX
PT   Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT   the proteins, useful as vaccines for raising immune response in animals.
XX
PS   Claim 7; Page 115; 127pp; English.
XX
CC   The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC   peptides derived from them. The invention is useful for producing an
CC   immune response. It is useful for preventing, treating or ameliorating a
CC   disorder or disease associated with infection of an animal with
CC   Helicobacter. Pharmaceutical composition and vaccines comprising the
CC   sequences of the invention is useful for treating type B gastritis,
CC   peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC   cell lymphoma. The invention is useful as reagents for clinical or
CC   medical diagnosis of Helicobacter infections and for scientific research
CC   on the properties of pathogenicity, virulence and infectivity of
CC   Helicobacter, as well as host defense mechanisms. The present sequence is
CC   Helicobacter sp. HP30-derived peptide
XX
SQ   Sequence 38 AA;

Alignment Scores:
Pred. No.:                2.34e-13        Length:          38
Score:                    198.00          Matches:         38
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              14.68%          Indels:          0
DB:                       5               Gaps:            0

US-09-732-091-3 (1-759) x AAE26873 (1-38)

Qy        157 GATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AspTyrAlaLysTyrAlaGluArgIleAlaGluGluLeuGlnTyrTyrGlySerAsnSer 20

Qy        217 TTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGAT 270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaSerPheIleLysGlyGluGlyValLeuTyrLysGluIleLeuCysAsp 38

RESULT 11
AAE26875
ID   AAE26875 standard; peptide; 41 AA.
XX
AC   AAE26875;
XX
```

DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30-derived peptide #4.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44528.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Claim 7; Page 116; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
SQ Sequence 41 AA;

Alignment Scores:

Pred. No.:	6.82e-13	Length:	41
Score:	194.00	Matches:	41
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	14.38%	Indels:	0
DB:	5	Gaps:	0

US-09-732-091-3 (1-759) x AAE26875 (1-41)

```
Qy        433 AACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCT 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGlyGlyPheLysSer 20

Qy        493 TATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 TyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeuGlyArgGlyLeu 40

Qy        553 TCG 555
              |||
Db         41 Ser 41
```

RESULT 12
AAE26874
ID AAE26874 standard; peptide; 30 AA.
XX
AC AAE26874;
XX
DT 13-DEC-2002 (first entry)
XX
DE Helicobacter sp. HP30-derived peptide #3.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44527.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Claim 7; Page 115; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
SQ Sequence 30 AA;

Alignment Scores:

Pred. No.:	3.27e-08	Length:	30
Score:	152.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	11.27%	Indels:	0
DB:	5	Gaps:	0

US-09-732-091-3 (1-759) x AAE26874 (1-30)

```
Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 20

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGC 450
              ||||||||||||||||||||||||||||||
Db         21 ThrAspAsnLeuAsnArgGlnAlaLeuSer 30
```

RESULT 13
AAE26872
ID AAE26872 standard; peptide; 30 AA.
XX
AC AAE26872;
XX
DT 13-DEC-2002 (first entry)
XX

DE Helicobacter sp. HP30-derived peptide #1.
XX
KW HP30; HP56; immune response; therapy; Helicobacter infection; vaccine;
KW type B gastritis; antiinflammatory; adenocarcinoma; defense mechanism;
KW low grade B cell lymphoma; virulence; antibacterial; gastric cancer;
KW immunostimulant; cytostatic; peptic ulcer.
XX
OS Helicobacter sp.
XX
PN WO200251237-A2.
XX
PD 04-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US048392.
XX
PR 07-DEC-2000; 2000US-00732091.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tian J, Walker R, Jackson WJ;
XX
DR WPI; 2002-666854/71.
DR N-PSDB; AAD44525.
XX
PT Novel Helicobacter proteins, HP30 and HP56, and nucleic acids encoding
PT the proteins, useful as vaccines for raising immune response in animals.
XX
PS Claim 7; Page 115; 127pp; English.
XX
CC The invention relates to Helicobacter HP30 or HP56 polypeptide and
CC peptides derived from them. The invention is useful for producing an
CC immune response. It is useful for preventing, treating or ameliorating a
CC disorder or disease associated with infection of an animal with
CC Helicobacter. Pharmaceutical composition and vaccines comprising the
CC sequences of the invention is useful for treating type B gastritis,
CC peptic ulcers, gastric cancers such as adenocarcinoma, and low grade B
CC cell lymphoma. The invention is useful as reagents for clinical or
CC medical diagnosis of Helicobacter infections and for scientific research
CC on the properties of pathogenicity, virulence and infectivity of
CC Helicobacter, as well as host defense mechanisms. The present sequence is
CC Helicobacter sp. HP30-derived peptide
XX
SQ Sequence 30 AA;

Alignment Scores:

Pred. No.:	5.5e-08	Length:	30
Score:	150.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	11.12%	Indels:	0
DB:	5	Gaps:	0

US-09-732-091-3 (1-759) x AAE26872 (1-30)

```
Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGT 90
              ||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGly 30
```

RESULT 14
AAW44365
ID AAW44365 standard; peptide; 173 AA.
XX
AC AAW44365;
XX
DT 28-MAY-1998 (first entry)
XX
DE Breast cancer-associated protein fragment BC-2 SEQ ID NO:12.
XX
KW Human; breast cancer-associated protein; nuclear matrix protein;
KW detection; diagnosis; antibody.
XX
OS Homo sapiens.
XX
PN WO9746884-A1.
XX
PD 11-DEC-1997.
XX
PF 03-JUN-1997; 97WO-US009529.
XX
PR 05-JUN-1996; 96US-00658639.
XX
PA (MATR-) MATRITECH INC.
XX
PI Keesee SK, Obar R, Wu Y;
XX
DR WPI; 1998-042336/04.
DR N-PSDB; AAV15304.
XX
PT Diagnosing breast cancer by detecting a breast cancer-associated protein
PT - allows early and reliable diagnosis and treatment monitoring, and
PT antibody or inhibitory compounds useful for treating breast cancer.
XX
PS Claim 11; Page 32; 47pp; English.
XX
CC The present sequence represents a breast cancer-associated protein
CC fragment for use in a method for diagnosing breast cancer in tissue or
CC body fluid by detecting one or more breast cancer-associated protein(s).
CC Alternatively a nucleic acid encoding a breast cancer- associated protein
CC is detected in the sample by the use of a nucleic acid probe. The breast
CC cancer-associated proteins, encoding nucleic acids and antibodies are
CC useful in diagnostic assays and kits for breast cancer detection. The
CC proteins are also useful in screening for inhibitory compounds and
CC monitoring effectiveness of treatments. The antibody, or pharmaceutical
CC compositions containing the antibody or the inhibitory compound, can be
CC administered to patients to treat breast cancer. The methods allow
CC reliable and early diagnosis of breast cancer, or prediction of its
CC onset, by detection of specific markers expressed in breast tumour cells
CC but not in non-cancerous breast cells
XX
SQ Sequence 173 AA;

Alignment Scores:

Pred. No.:	0.0325	Length:	173
Score:	101.50	Matches:	41
Percent Similarity:	43.58%	Conservative:	37
Best Local Similarity:	22.91%	Mismatches:	72
Query Match:	7.52%	Indels:	29
DB:	2	Gaps:	5

US-09-732-091-3 (1-759) x AAW44365 (1-173)

```
Qy         26 TGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTTT 85
              |||         :::  |||  |||        |||
Db          2 TrpProSerLysAlaArgTrpMetLeuPheAlaSerTrp---------------------- 14

Qy         86 TTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATACAAAA 145
                                                |||      ||||||       :::
Db         15 --------------------------GlnLysThrTrpValAlaProGlyTyrValArg 25

Qy        146 GGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATG 205
                  ::::::      |||  |||      |||||||||  :::|||
Db         26 LysPheValLeuMetArgAlaAsnIleGlnAlaValSerLeuLys--IleGlnThrLeuL 45

Qy        206 GGAGC---AATAGTTTTGCGAGTTTCATTAAAGGC------------GAAGGAGTCTTAT 250
                |||   ||||||  |||      :::||||||               |||   :::
Db         45 ysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAlaMetGlyThrMetA 65

Qy        251 ACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTG 310
                 :::::::::                 |||:::  :::  :::  ::::::  |
Db         65 snArgGlnLeuLysLeuProGlnIleGlnLysIleMetMetGluPheGluArgGlnAlaG 85

Qy        311 AAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAA 370
```

```
Db        85 luIleMetAspMetLysGluGluMetMetAsnAspAlaIleAspAspAlaMetGlyAspG 105

Qy       371 TGGATGATGAAGAA--------------GTGAAAGAAATGTGCGATGAATTATCCATAA 415
Db       105 luGluAspGluGluGluSerAspAlaValValSerGlnValLeuAspGluLeuGlyLeuS 125

Qy       416 AAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAA 475
Db       125 erLeuThrAspGluLeuSerAsnLeuProSerThrGlyGlySerLeuSerVal---AlaA 144

Qy       476 TGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCA 528
Db       144 laGlyGlyLysLysAlaGluAlaAlaAlaSerAlaLeuAlaAspAlaAspAla 161

RESULT 15
AAB98717
ID   AAB98717 standard; protein; 173 AA.
XX
AC   AAB98717;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Human breast cancer-associated protein BC-2 portion, SEQ ID NO:12.
XX
KW   Human breast cancer-associated protein; BC-2; marker protein;
KW   nuclear matrix protein; diagnosis; detection; tumour.
XX
OS   Homo sapiens.
XX
PN   US6218131-B1.
XX
PD   17-APR-2001.
XX
PF   06-OCT-1997;   97US-00944604.
XX
PR   05-JUN-1996;   96US-00658639.
XX
PA   (MATR-) MATRITECH INC.
XX
PI   Keesee SK,  Obar R,  Wu Y;
XX
DR   WPI; 2001-396355/42.
DR   N-PSDB; AAH25091.
XX
PT   Diagnosing or detecting breast cancer in an individual comprises
PT   detecting the presence of breast cancer-associated proteins in a
PT   biological sample.
XX
PS   Example 1; Col 21-24; 30pp; English.
XX
CC   The invention relates to novel human breast cancer-associated proteins,
CC   and their use in diagnosing and detecting breast cancer. The breast
CC   cancer-associated proteins of the invention are nuclear matrix proteins
CC   designated BC-2 (AAB98720), BC-8 isoform A (BC-8A, AAB98721) and CC BC-8
CC   isoform B (BC-8B, AAB98722), although 6 other breast cancer-associated
CC   nuclear matrix proteins (BC-1, BC-3, BC-4, BC-5, BC-6, and BC-7) were
CC   also isolated (sequences not given in the specification). The novel
CC   breast cancer-associated proteins are present in the nuclear matrix of
CC   breast cancer cells, but are not present in the nuclear matrix of cells
CC   from normal breast tissue. The invention also encompasses fragments of
CC   the breast cancer-associated proteins (AAB98709-AAB98716), and methods
CC   for their use in breast cancer diagnosis. The breast cancer marker
CC   proteins of the invention or fragments thereof may be detected in a
CC   sample of breast tissue or blood from an individual. Isolated breast
CC   cancer-associated proteins may also be used to prepare antibodies,
CC   particularly monoclonal antibodies, which may be used to detect or
CC   diagnose breast cancer. Nucleic acids encoding the breast cancer-
CC   associated proteins may also be used in the diagnosis and detection of
CC   breast cancer, and in the isolation of DNA or protein sequences which may
CC   interact with breast cancer-associated nuclear matrix proteins. The
CC   present sequence represents a portion of breast cancer-associated protein
CC   BC-2 referred to in an exemplification of the invention
XX
SQ   Sequence 173 AA;

Alignment Scores:
Pred. No.:                0.0325          Length:          173
Score:                    101.50          Matches:         41
Percent Similarity:       43.58%          Conservative:    37
Best Local Similarity:    22.91%          Mismatches:      72
Query Match:              7.52%           Indels:          29
DB:                       4               Gaps:            5

US-09-732-091-3 (1-759) x AAB98717 (1-173)

Qy        26 TGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTTT 85
Db         2 TrpProSerLysAlaArgTrpMetLeuPheAlaSerTrp---------------------- 14

Qy        86 TTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATACAAAA 145
Db        15 --------------------------GlnLysThrTrpValAlaProGlyTyrValArg 25

Qy       146 GGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATG 205
Db        26 LysPheValLeuMetArgAlaAsnIleGlnAlaValSerLeuLys--IleGlnThrLeuL 45

Qy       206 GGAGC---AATAGTTTTGCGAGTTTCATTAAAGGC-----------GAAGGAGTCTTAT 250
Db        45 ysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAlaMetGlyThrMetA 65

Qy       251 ACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTG 310
Db        65 snArgGlnLeuLysLeuProGlnIleGlnLysIleMetMetGluPheGluArgGlnAlaG 85

Qy       311 AAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAA 370
Db        85 luIleMetAspMetLysGluGluMetMetAsnAspAlaIleAspAspAlaMetGlyAspG 105

Qy       371 TGGATGATGAAGAA--------------GTGAAAGAAATGTGCGATGAATTATCCATAA 415
Db       105 luGluAspGluGluGluSerAspAlaValValSerGlnValLeuAspGluLeuGlyLeuS 125

Qy       416 AAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAA 475
Db       125 erLeuThrAspGluLeuSerAsnLeuProSerThrGlyGlySerLeuSerVal---AlaA 144

Qy       476 TGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCA 528
Db       144 laGlyGlyLysLysAlaGluAlaAlaAlaSerAlaLeuAlaAspAlaAspAla 161

Search completed: July  5, 2004, 03:03:13
Job time : 69 secs
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 5, 2004, 03:00:58 ; Search time 22.5 Seconds
(without alignments)
6489.722 Million cell updates/sec

Title: US-09-732-091-3
Perfect score: 1349
Sequence: 1 atggcatacaaatatgatag.....cgttgcaaatagaatccatt 759

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09732091/runat_02072004_182407_19036/app_query.fasta_1.903
-DB=PIR_78 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09732091_@CGN_1_1_25_@runat_02072004_182407_19036 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1279	94.8	253	2	D64718	conserved hypothet
	2	1270	94.1	253	2	B71800	hypothetical prote
	3	784.5	58.2	207	2	E64718	conserved hypothet
	4	630	46.7	209	2	A71800	hypothetical prote
	5	472.5	35.0	155	2	C64718	conserved hypothet
	6	265	19.6	237	2	C64721	hypothetical prote
	7	265	19.6	237	2	AC0503	conserved hypothet
	8	262	19.4	237	2	C90630	probable oxidoredu
	9	262	19.4	237	2	C85481	probable oxidoredu
	10	133	9.9	39	2	F64718	hypothetical prote
c	11	127.5	9.5	206	2	D90630	hypothetical prote
c	12	127.5	9.5	206	2	D85481	hypothetical prote
c	13	122	9.1	196	2	A40623	heat shock protein
	14	98.5	7.3	1009	2	C89910	hypothetical prote
	15	98	7.3	282	2	AE0360	hypothetical prote
	16	98	7.3	1099	2	G90546	conserved hypothet
	17	97.5	7.2	311	2	C81380	probable D-2-hydro
	18	97.5	7.2	728	2	A81385	probable ATP /GTP
c	19	97	7.2	346	2	A70144	hypothetical prote
	20	97	7.2	916	2	E71330	probable preprotei
	21	96.5	7.2	373	2	G70355	8-amino-7-oxononan
	22	96	7.1	262	2	T33071	hypothetical prote
	23	96	7.1	420	2	F96920	probable permease
	24	95.5	7.1	344	2	A47025	abortive phage res
	25	95	7.0	517	2	C81327	hypothetical prote
	26	95	7.0	2649	2	A40937	bullous pemphigoid
	27	94.5	7.0	387	2	A86302	hypothetical prote
	28	93.5	6.9	802	2	S49252	penicillin amidase
	29	93	6.9	355	2	E90488	transposase ISC122
	30	93	6.9	610	2	G70178	exodeoxyribonuclea
	31	92.5	6.9	430	2	T33070	hypothetical prote
	32	92.5	6.9	802	2	I39665	penicillin amidase
	33	92.5	6.9	886	2	T16536	hypothetical prote
	34	92	6.8	308	2	H81345	hypothetical prote
	35	92	6.8	1939	2	T18372	repeat organellar
	36	91.5	6.8	399	2	G72253	RNA polymerase sig
	37	91.5	6.8	472	2	A64320	PET112 homolog - M
	38	91.5	6.8	508	2	T24622	hypothetical prote
	39	91.5	6.8	615	2	S42797	rad 26 protein - f
	40	91.5	6.8	1769	2	S53378	probable membrane
	41	91.5	6.8	1847	2	E64477	replication factor
	42	91	6.7	284	2	A45488	body-wall muscle t
	43	91	6.7	359	2	AI2368	hypothetical prote
	44	90.5	6.7	320	2	B97214	hypothetical prote
	45	90.5	6.7	440	2	E64571	transposase-like p

ALIGNMENTS

RESULT 1
D64718
conserved hypothetical protein HP1588 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: D64718
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: D64718
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-253 <TOM>
A;Cross-references: GB:AE000656; GB:AE000511; NID:g2314771; PIDN:AAD08627.1; PID:g2314

Alignment Scores:

Pred. No.:	1.5e-97	Length:	253
Score:	1279.00	Matches:	253
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	94.81%	Indels:	0
DB:	2	Gaps:	0

US-09-732-091-3 (1-759) x D64718 (1-253)

```
Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
```

```
Db        41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy       181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240

Db        61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy       241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300

Db        81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy       301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360

Db       101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy       361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420

Db       121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy       421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480

Db       141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy       481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540

Db       161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy       541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600

Db       181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy       601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660

Db       201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy       661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720

Db       221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy       721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759

Db       241 AlaAsnGlyAspLysLysSerLeuGlnIleGluSerIle 253
```

RESULT 2
B71800
hypothetical protein jhp1494 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C;Accession: B71800
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: B71800
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <ARN>
A;Cross-references: GB:AE001571; GB:AE001439; NID:g4156120; PIDN:AAD07073.1; PID:g415612
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp1494

Alignment Scores:

Pred. No.:	8.25e-97	Length:	253
Score:	1270.00	Matches:	251
Percent Similarity:	99.60%	Conservative:	1
Best Local Similarity:	99.21%	Mismatches:	1
Query Match:	94.14%	Indels:	0
DB:	2	Gaps:	0

US-09-732-091-3 (1-759) x B71800 (1-253)

```
Qy         1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60

Db         1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy        61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120

Db        21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy       121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180

Db        41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy       181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240

Db        61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy       241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300

Db        81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy       301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360

Db       101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy       361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420

Db       121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy       421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480

Db       141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy       481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540

Db       161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy       541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600

Db       181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy       601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660

Db       201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy       661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720

Db       221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy       721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759

Db       241 AlaAsnGluAspLysLysSerLeuGlnIleGluSerVal 253
```

RESULT 3
E64718
conserved hypothetical protein HP1589 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Aug-1997
C;Accession: E64718
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: E64718
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-207 <TOM>
A;Cross-references: GB:AE000511; TIGR:HP1589
C;Genetics:
A;Start codon: TTG

Alignment Scores:

Pred. No.:	8.01e-57	Length:	207
Score:	784.50	Matches:	153
Percent Similarity:	83.65%	Conservative:	21
Best Local Similarity:	73.56%	Mismatches:	33
Query Match:	58.15%	Indels:	1
DB:	2	Gaps:	1

US-09-732-091-3 (1-759) x E64718 (1-207)

```
Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              :::|||||||   |||||:::|||:::|||   |||||||||||||||      |||
Db          1 MetThrSerSerThrGluTyrGlnArgTyrGlyTyrAspTyrAlaLysTyrProArgArg 20

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||   ||||||  ||||||||  :::|||  :::   |||
Db         21 IleAlaGluGluLeuGlnArgTyrGlyGlyAsnSerPheMetAsnPhePheArgAspGlu 40

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||  ||||||  ||||||||||||||||||
Db         41 GlyValLeuTyrLysGluIleLeuCysAspAlaCysAspHisLeuLysValAsnTyrAsn 60

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||::::::|||||||||||||||:::||||||||||||  ||||||::::::|||
Db         61 LysLysSerAspThrThrLeuIleGluGluAsnMetLeuSerSerIleLeuGlnLysSer 80

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||:::|||   ||||||||:::::::|||:::||||||||||||   :::||||||
Db         81 LeuGluLysMetSerAspGluGluIleArgGluLeuCysAspGluLeuGlyValLysAsn 100

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              |||:::   |||   :::|||||||||||   |||   ||||||||||:::||||||
Db        101 ThrAsnLysLeuGlyLysGlnAlaLeuSerThrAlaAlaLeuThrLeuPheArgMetGly 120

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||:::|||||||||||||||||||   |||   |||
Db        121 GlyPheLysSerTyrGlnLeuAlaLeuIleValAlaAsnAlaValIleLysAlaIlePhe 140

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              |:|||||||||||||||   |||   ||||||||  ||||||  ||||||
Db        141 GlnArgGlyLeuSerLeuGlyAlaAsnAlaAlaLeuThrArgGlyLeuSerIleLeuThr 160

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyProIleGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 180

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||:::|||||||||||||||||||||      |||
Db        181 TyrArgValThrIleProAlaCysIleLeuValAlaThrLeuArgLeuLysAla---Gln 199

Qy        721 GCCAATGGAGATAAGAAGTCGTTG 744
              ||||||   |||   |||
Db        200 AlaAsnGluIleLysAsnIleLeu 207
```

RESULT 4
A71800
hypothetical protein jhp1493 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C;Accession: A71800
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: A71800
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-209 <ARN>
A;Cross-references: GB:AE001571; GB:AE001439; NID:g4156120; PIDN:AAD07072.1; PID:g415612
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp1493

Alignment Scores:

Pred. No.:	4.25e-44	Length:	209
Score:	630.00	Matches:	122
Percent Similarity:	75.49%	Conservative:	32
Best Local Similarity:	59.80%	Mismatches:	46
Query Match:	46.70%	Indels:	4
DB:	2	Gaps:	2

US-09-732-091-3 (1-759) x A71800 (1-209)

```
Qy        112 AATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATAC 171
              ||||||:::|||||||||      ||||||:::|||:::|||   ||||||||||||||
Db          2 AsnGluGluLeuThrSerLeuThrGluTyrGlnArgTyrGlyHisAspTyrAlaLysTyr 21

Qy        172 GCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATT 231
              ||||||||||||||||||||||   ||||||   ||||||||||||:::|||
Db         22 ProArgArgIleAlaGluGluLeuGlnArgTyrGlyGlyAsnSerPheAlaAsnPhePhe 41

Qy        232 AAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTC 291
              :::   ||||||||||||||||||||||||||||||||   ||||||   |||   :::
Db         42 ArgAspGluGlyValLeuTyrLysGluIleLeuCysAspAlaCysAspHisLeuAspIle 61

Qy        292 AATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTA 351
              |||||||||:::::::::   |||:::|||||||||||||||||||||||||:::|||
Db         62 AsnTyrAsnGluArgSerAlaThrSerLeuIleGluGlnAsnMetLeuSerLysLeuLeu 81

Qy        352 GAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCC 411
              :::   |||||||||:::|||          |||:::|||||||:::|||||   |||
Db         82 LysAspSerLeuGluLysMetSerGlyArgGluIleLysGluLeuCysAspGlyLeuGly 101

Qy        412 ATAAAAAACACGGACAATTTA---------AACAGACAAGCCTTAAGCGCGGCGACTTTA 462
              :::   |||   |||   :::          |||:::|||   |||   |||:::   |||
Db        102 MetProAsnIleAspLysValIleGlyGluAsnLysGlnValLeuIleAlaSerValLeu 121

Qy        463 ACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCG 522
              ||||||||||||   ||||||      ||||||   |||||||||   ||||||:::|||
Db        122 ThrLeuPheLysAlaGlyGlySerHisSerTyrAlaLeuAlaValAlaValAlaAspAla 141

Qy        523 GTCGCAAAAACCATTCTAGGGCGTGGTTTA---TCGCTTGCGGGCAATCAGGTGCTTACA 579
              :::   :::       ||||||   ||||||   |||:::   |||           |||
Db        142 MetValArgGlnThrLeuGlyHisGlyLeuSerSerValValGlyLysValAlaLeuLys 161

Qy        580 AGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCG 639
              :::||||||       |||   ||||||:::||||||:::||||||||||      :::
Db        162 LysThrLeuAspIleLeuAlaGlyProIleGlyTrpValIleThrGlyAlaLeuValSer 181

Qy        640 ATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACT 699
              |||::::::||||||||||||||||||||||||:::|||||||||::::::||||||||
Db        182 IleAsnLeuAlaGlyProAlaTyrArgValThrValProAlaCysValLeuValAlaThr 201

Qy        700 TTACGCCTAAAA 711
              ||||||   |||
Db        202 LeuArgLysLys 205
```

RESULT 5
C64718
conserved hypothetical protein HP1587 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: C64718
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R. Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKer son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: C64718
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

A;Residues: 1-155 <TOM>
A;Cross-references: GB:AE000656; GB:AE000511; NID:g2314771; PIDN:AAD08626.1; PID:g231477

Alignment Scores:

Pred. No.:	3.99e-31	Length:	155
Score:	472.50	Matches:	93
Percent Similarity:	78.23%	Conservative:	22
Best Local Similarity:	63.27%	Mismatches:	29
Query Match:	35.03%	Indels:	3
DB:	2	Gaps:	1

US-09-732-091-3 (1-759) x C64718 (1-155)

```
Qy        112 AATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATAC 171
              ||||||  ||||||:::|||  ||||||||||||:::||  ||||||||||||||||
Db          2 AsnGluAspLeuThrAsnSerThrGluTyrLysArgTyrGlyHisAspTyrAlaLysTyr 21

Qy        172 GCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATT 231
              ||||||||||||||||||||||:::||||||  ||||||||||:::|||
Db         22 ProArgArgIleAlaGluGluLeuGlnHisTyrGlyGlyAsnSerPheAlaAsnPhePhe 41

Qy        232 AAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTC 291
              :::  ||||||||||||||||||||||||||||||||  ||||||  ||||||||
Db         42 ArgAspGluGlyValLeuTyrLysGluIleLeuCysAspAlaCysAspHisLeuLysVal 61

Qy        292 AATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTA 351
              |||||||||:::::::::  |||:::||||||||||||||||||||||||||:::|||
Db         62 AsnTyrAsnGluGluSerAlaThrSerLeuIleGluGlnAsnMetLeuSerLysLeuLeu 81

Qy        352 GAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCC 411
              :::  |||||||||:::|||          |||:::||||||:::|||:::||||||
Db         82 LysAspSerLeuGluLysMetSerArgArgGluIleLysGluLeuCysAsnGluLeuGly 101

Qy        412 ATAAAAAACACGGACAATTTA---------AACAGACAAGCCTTAAGCGCGGCGACTTTA 462
              :::  |||  |||  :::          |||:::|||  |||  |||:::|||||
Db        102 MetThrAsnIleAspLysValIleGlyGluAsnLysGlnValLeuIleAlaSerThrLeu 121

Qy        463 ACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCG 522
              ||||||||||||  ||||||  ||||||  |||||||||  ||||||:::|||
Db        122 ThrLeuPheLysAlaGlyGlySerHisSerTyrAlaLeuAlaValSerValAlaAspAla 141

Qy        523 GTCGCAAAAACCATTCTAGGG 543
              :::  :::  ||||||
Db        142 MetValArgGlnThrLeuGly 148
```

RESULT 6
C64721
hypothetical protein b0011 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: C64721; D56688; S28462
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: C64721
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-237 <BLAT>
A;Cross-references: GB:AE000212; GB:U00096; NID:g1786192; PIDN:AAC73122.1; PID:g1786193;
A;Experimental source: strain K-12, substrain MG1655
R;James, R.; Dean, D.O.; Debbage, J.
DNA Seq. 3, 327-332, 1993
A;Title: Five open reading frames upstream of the dnaK gene of Escherichia coli.
A;Reference number: A56688; MUID:94003405; PMID:8400364
A;Accession: D56688
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-189,'F',191-237 <JAM>
A;Cross-references: GB:X67700; NID:g41754; PIDN:CAA47934.1; PID:g41759

Alignment Scores:

Pred. No.:	4.92e-14	Length:	237
Score:	265.00	Matches:	77
Percent Similarity:	48.99%	Conservative:	44
Best Local Similarity:	31.17%	Mismatches:	90
Query Match:	19.64%	Indels:	36
DB:	2	Gaps:	6

US-09-732-091-3 (1-759) x C64721 (1-237)

```
Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
              |||     |||  ||||||:::||||||::                |||   :::
Db          5 TyrLeuAsnAspSerAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeuAlaAsn 24

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
                    :::|||          :::  |||:::  |||  :::      |||
Db         25 PheAlaArgLeuLeuThrHisAsnGluLysGlyLysThrArgLeuSerSerValLeuMet 44

Qy        127 AGCTCCATAGAATACAAAAGG---------CATGGCGATGATTACGCTAAATACGCAGAA 177
                 :::      :::|||             |||  :::  :::  :::      :::
Db         45 ArgAsnGluLeuPheLysSerMetGluGlyHisProGluGlnHisArgArgAsnTrpGln 64

Qy        178 AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGC 237
              ||||||  ||||||||:::::|||  :::|||  |||:::  ::::::|||
Db         65 LeuIleAlaGlyGluLeuGlnHisPheGlyGlyAspSerIleAlaAsnLysLeuArgGly 84

Qy        238 GAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTAC 297
              |||  ||||||:::  ||||||  ||||||      :::|||||:::
Db         85 HisGlyLysLeuTyrArgAlaIleLeuLeuAspValSerLysArgLeuLysLeuLysAla 104

Qy        298 AACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGA 357
              :::|||:::      |||      ||||||||  :::|||  :::  |||
Db        105 AspLysGluMetSerThrPheGluIleGluGlnGlnLeuLeuGluGlnPheLeuArgAsn 124

Qy        358 AGTTTGGAAGAAATGGATGATGAA--------------------------------GAA 384
              :::  ::::::||||||:::|||
Db        125 ThrTrpLysLysMetAspGluGluHisLysGlnGluPheLeuHisAlaValAspAlaArg 144

Qy        385 GTGAAAGAAATGTGCGATGAATTATCCATAAAAAAACACGGACAATTTAAACAGACAAGCC 444
              |||  |||:::  :::  |||  :::          |||  |||      :::
Db        145 ValAsnGluLeuGluGluLeuProLeuLeuMetLysAspLysLeuAlaLysGly 164

Qy        445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCT 504
              :::|||                                  |||  ||||||
Db        165 ValSer-----------------------------HisLeuLeuSerSerGlnLeuThr 174

Qy        505 GTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGC 564
              |||:::              |||  :::::||||||  ||||||  |||  |||
Db        175 ArgIleLeuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu---LeuArgGly 193

Qy        565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACA 624
                                  :::  |||  ||||||||||||      :::
Db        194 ------------------------AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205

Qy        625 GGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGC 684
              ||||||                :::::|||  ||||||||||||||||||||||||
Db        206 GlyValLysAla---------ValSerGlySerAlaTyrArgValThrIleProAlaVal 222

Qy        685 ATTGTGGTTGCCACTTTACGC 705
              :::  :::|||  ||||||
Db        223 LeuGlnIleAlaCysLeuArg 229
```

RESULT 7
AC0503
conserved hypothetical protein STY0010 [imported] - Salmonella enterica subsp. enteri
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0503
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11577608
A;Accession: AC0503
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-237 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD01163.1; PID:g16501293; GSPDB:GN00176
C;Genetics:
A;Gene: STY0010

Alignment Scores:

Pred. No.:	4.92e-14	Length:	237
Score:	265.00	Matches:	79
Percent Similarity:	49.17%	Conservative:	40
Best Local Similarity:	32.64%	Mismatches:	101
Query Match:	19.64%	Indels:	22
DB:	2	Gaps:	7

US-09-732-091-3 (1-759) x AC0503 (1-237)

```
Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              :::   |||   :::|||   |||||:::||||||:::                 |||
Db          3 ValThrTyrLeuHisAspGluAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeu 22

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGA---------- 108
                |||       :::|||         :::   |||:::   |||
Db         23 AlaAspPheAlaArgLeuLeuThrHisAsnGluLysGlyLysAlaArgLeuSerSerVal 42

Qy        109 ------CACAATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTAC 162
                   |||||||||   |||      ::::::|||         |||   :::   :::
Db         43 LeuSerHisAsnGlu---LeuPheLysAlaMetGlu------GlyHisProGluGlnHis 59

Qy        163 GCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCG 222
                 :::      :::   ||||||   |||   |||:::|||||   :::|||   |||
Db         60 ArgArgAsnTrpGlnLeuIleAlaGlyGluPheGlnHisTyrGlyGlyAspSerIleAla 79

Qy        223 AGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAA 282
              :::   ::::::|||   |||      |||:::   ||||||   ||||||      :::
Db         80 AsnLysLeuArgGlyHisGlyLysGlnTyrArgAlaIleLeuLeuAspValAlaLysArg 99

Qy        283 TTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCT 342
              ||||||:::      :::|||         |||      |||||||||   :::|||
Db        100 LeuLysLeuLysAlaAspLysSerMetSerThrPheGluIleGluGlnGlnLeuLeuGlu 119

Qy        343 AAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGAT 402
                 |||      :::   ::::::||||||            :::|||
Db        120 HisPheLeuArgHisThrTrpGlnLysMetAspAlaAlaHisLysGlnGluPheLeuGln 139

Qy        403 GAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTA 462
              :::      |||   ::::::          |||
Db        140 AlaValAspAlaLysValSerGluLeuGluGluLeuLeuProLeuLeuMetLysAspArg 159

Qy        463 ACGCTGTTTAAA---ATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAAT 519
              :::|||   |||   :::          |||   ||||||      |||:::
Db        160 SerLeuAlaLysGlyValSerHisLeuLeuSerThrGlnLeuThrArgIleLeuArgThr 179

Qy        520 GCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACA 579
                 |||   :::|||||||||   ||||||   |||   |||
Db        180 HisAlaAlaMetSerIleLeuGlyHisGlyLeu---LeuArgGly--------------- 193

Qy        580 AGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCG 639
                    :::   |||   ||||||||||||      :::   ||||||
Db        194 ---------AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsnGlyValLysAla--- 209

Qy        640 ATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACT 699
                  :::::::|||   |||||||||||||||||||||||||   :::   :::|||
Db        210 ------MetSerGlySerAlaTyrArgValThrIleProAlaValLeuGlnIleAlaCys 227

Qy        700 TTACGC 705
                 ||||||
Db        228 LeuArg 229
```

RESULT 8

C90630
probable oxidoreductase ECs0012 [imported] - Escherichia coli (strain O157:H7, substra
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: C90630
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ge
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90630
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-237 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33434.1; PID:g13359467; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0012

Alignment Scores:

Pred. No.:	8.69e-14	Length:	237
Score:	262.00	Matches:	76
Percent Similarity:	48.99%	Conservative:	45
Best Local Similarity:	30.77%	Mismatches:	90
Query Match:	19.42%	Indels:	36
DB:	2	Gaps:	6

US-09-732-091-3 (1-759) x C90630 (1-237)

```
Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
              |||      |||   ||||||:::||||||:::                  |||   :::
Db          5 TyrLeuAsnAspSerAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeuAlaAsn 24

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
                  :::|||         :::   |||:::   |||   :::      |||
Db         25 PheAlaArgLeuLeuThrHisAsnGluLysGlyLysThrArgLeuSerSerValLeuMet 44

Qy        127 AGCTCCATAGAATACAAAAGG---------CATGGCGATGATTACGCTAAATACGCAGAA 177
                 :::      :::|||            |||   :::   :::   :::      :::
Db         45 ArgAsnGluLeuPheLysSerMetGluGlyHisProGluGlnHisArgArgAsnTrpGln 64

Qy        178 AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGC 237
                 ||||||   |||||||||:::::::|||   :::|||   |||:::   ::::::|||
Db         65 LeuIleAlaGlyGluLeuGlnHisPheGlyGlyAspSerIleAlaAsnLysLeuArgGly 84

Qy        238 GAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTAC 297
                 |||   ||||||:::   ||||||   ||||||      :::||||||:::
Db         85 HisGlyLysLeuTyrArgAlaIleLeuLeuAspValSerLysArgLeuLysLeuLysAla 104

Qy        298 AACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGA 357
              :::|||:::      |||      |||||||||   :::|||   :::   |||
Db        105 AspLysGluMetSerThrPheGluIleGluGlnGlnLeuLeuGluGlnPheLeuArgAsn 124

Qy        358 AGTTTGGAAGAAATGGATGATGAA--------------------------------GAA 384
              :::   ::::::|||||||:::|||
Db        125 ThrTrpLysLysMetAspGluGluHisLysGlnGluPheLeuHisAlaValAspAlaArg 144

Qy        385 GTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCC 444
              |||   |||:::   :::   |||   :::      |||   |||      :::
Db        145 ValAsnGluLeuGluGluLeuLeuProLeuLeuMetLysAspLysLeuLeuAlaLysGly 164

Qy        445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCT 504
              :::|||                                  |||   ||||||
Db        165 ValSer--------------------------------HisLeuLeuSerSerGlnLeuThr 174

Qy        505 GTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGC 564
                 |||:::            |||   ::::::||||||   ||||||   |||   |||
Db        175 ArgIleLeuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu---LeuArgGly 193

Qy        565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACA 624
                                    :::   |||   ||||||||||||      :::
Db        194 ------------------------AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205
```

```
Qy        625 GGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGC 684
              ||||||                 ::::::|||   :::||||||||||||||||||||
Db        206 GlyValLysAla---------ValSerGlySerSerTyrArgValThrIleProAlaVal 222

Qy        685 ATTGTGGTTGCCACTTTACGC 705
              :::   :::|||   ||||||
Db        223 LeuGlnIleAlaCysLeuArg 229
```

RESULT 9
C85481
probable oxidoreductase Z0011 [imported] - Escherichia coli (strain O157:H7, substrain E
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: C85481
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85481
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-237 <STO>
A;Cross-references: GB:AE005174; NID:g12512689; PIDN:AAG54311.1; GSPDB:GN00145; UWGP:Z00
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0011

Alignment Scores:

Pred. No.:	8.69e-14	Length:	237
Score:	262.00	Matches:	76
Percent Similarity:	48.99%	Conservative:	45
Best Local Similarity:	30.77%	Mismatches:	90
Query Match:	19.42%	Indels:	36
DB:	2	Gaps:	6

US-09-732-091-3 (1-759) x C85481 (1-237)

```
Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
              |||      |||   ||||||:::||||||:::                |||   :::
Db          5 TyrLeuAsnAspSerAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeuAlaAsn 24

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
                  :::|||         :::   |||:::   |||   :::       |||
Db         25 PheAlaArgLeuLeuThrHisAsnGluLysGlyLysThrArgLeuSerSerValLeuMet 44

Qy        127 AGCTCCATAGAATACAAAAGG---------CATGGCGATGATTACGCTAAATACGCAGAA 177
                 :::      :::|||            |||   :::   :::   :::      :::
Db         45 ArgAsnGluLeuPheLysSerMetGluGlyHisProGluGlnHisArgArgAsnTrpGln 64

Qy        178 AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGC 237
              ||||||   |||||||||:::::::|||   :::|||   |||:::   :::::::|||
Db         65 LeuIleAlaGlyGluLeuGlnHisPheGlyGlyAspSerIleAlaAsnLysLeuArgGly 84

Qy        238 GAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTAC 297
                 |||   ||||||:::   ||||||   ||||||      :::||||||:::
Db         85 HisGlyLysLeuTyrArgAlaIleLeuLeuAspValSerLysArgLeuLysLeuLysAla 104

Qy        298 AACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGA 357
              :::|||:::      |||      ||||||||   :::|||   :::   |||
Db        105 AspLysGluMetSerThrPheGluIleGluGlnGlnLeuLeuGluGlnPheLeuArgAsn 124

Qy        358 AGTTTGGAAGAAATGGATGATGAA--------------------------------GAA 384
              :::   ::::::||||||:::|||
Db        125 ThrTrpLysLysMetAspGluGluHisLysGlnGluPheLeuHisAlaValAspAlaArg 144

Qy        385 GTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCC 444
              |||   |||:::   :::   |||   :::         |||   |||      :::
Db        145 ValAsnGluLeuGluGluLeuProLeuLeuMetLysAspLysLeuLeuAlaLysGly 164

Qy        445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCT 504
              :::|||                                   |||   ||||||
Db        165 ValSer-----------------------------HisLeuLeuSerSerGlnLeuThr 174
```

```
Qy        505 GTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGC 564
              |||:::            |||   ::::::||||||   ||||||   |||   |||
Db        175 ArgIleLeuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu---LeuArgGly 193

Qy        565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACA 624
                                    :::   |||   ||||||||||||      :::
Db        194 ----------------------AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205

Qy        625 GGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGC 684
              ||||||                 ::::::|||   :::||||||||||||||||||||
Db        206 GlyValLysAla---------ValSerGlySerSerTyrArgValThrIleProAlaVal 222

Qy        685 ATTGTGGTTGCCACTTTACGC 705
              :::   :::|||   ||||||
Db        223 LeuGlnIleAlaCysLeuArg 229
```

RESULT 10
F64718
hypothetical protein HP1590 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: F64718
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: F64718
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-39 <TOM>
A;Cross-references: GB:AE000656; GB:AE000511; NID:g2314771; PIDN:AAD08629.1; PID:g231

Alignment Scores:

Pred. No.:	0.00363	Length:	39
Score:	133.00	Matches:	26
Percent Similarity:	79.49%	Conservative:	5
Best Local Similarity:	66.67%	Mismatches:	8
Query Match:	9.86%	Indels:	0
DB:	2	Gaps:	0

US-09-732-091-3 (1-759) x F64718 (1-39)

```
Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              |||||||||:::||||||   ||||||||||||||||||:::|||   ||||||||||||
Db          1 MetAlaTyrArgTyrAspSerAspLeuGluPheLeuLysArgLeuSerSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAA 117
                 |||||||||:::   ||||||:::   :::||||||      |||   ||||||
Db         21 LysAspLeuPheAspAlaLeuValTyrAspGluAspGlyThrLeuArgMetAsnGlu 39
```

RESULT 11
D90630
hypothetical protein ECs0011 [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: D90630
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90630
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-206 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33435.1; PID:g13359468; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0011

C;Superfamily: Escherichia coli heat shock protein htgA

Alignment Scores:

Pred. No.:	0.0104	Length:	206
Score:	127.50	Matches:	63
Percent Similarity:	38.84%	Conservative:	24
Best Local Similarity:	28.12%	Mismatches:	96
Query Match:	9.46%	Indels:	41
DB:	2	Gaps:	7

US-09-732-091-3 (1-759) x D90630 (1-206)

```
Qy        629 ACGCCTGTAATGATCCAGCCAACAGGACCTGTTAAAAAGCTCAGAGTTCTTGTAAGCACC 570
              ||||||          ||||||||||||
Db          5 ThrProPheSerAlaAlaProThrGlyProProSer----------------------- 16

Qy        569 TGATTGCCCGCA---AGCGATAAACCACGCCCTAGAATGGTTTTTGCGACCGCATTCGCA 513
              |||||||         ||||||  |||      :::  |||
Db         17 ------ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaTrpValArg 34

Qy        512 ACAATGACAGCTAATTGATAAGATTTAAAA------CCCCCCATTTTAAACAGCGTTAAA 459
              |||      :::      :::  :::         |||      |||:::
Db         35 LysMetArgValSerTrpLeuGluSerArgCysAspThrProPheAlaAsnAsn------ 52

Qy        458 GTCGCCGCGCTTAAGGCTTGTCTGTTTAAATTGTCCGTGTTTTTTATGGATAATTCATCG 399
                                          ||||||                 :::||||||
Db         53 ----------------------------LeuSerPheIleSerSerGlySerSerSer 62

Qy        398 CACATTTCTTTCACT--------------------------------TCTTCATCATCC 372
              |||||||||                                      |||||||||
Db         63 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuCysSerSerSer 82

Qy        371 ATTTCTTCCAAACTTCTTTCTAAGATTTTAGAAAGCATGTTTTGTTCAATTAAAGTCGTT 312
              |||      ::::::|||   :::         |||      |||||||||      |||
Db         83 IlePhePheGlnValLeuArgArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 102

Qy        311 TCAGTTTTCTTGTTGTAATTGACCTTTAATTTATCGCACACATCGCATAAAATCTCTTTG 252
              :::  |||       :::||||||         ||||||  |||::::::
Db        103 AspIleSerLeuSerAlaPheSerPheAsnArgPheGluThrSerSerLysMetAlaArg 122

Qy        251 TATAAGACTCCTTCGCCTTTAATGAAACTCGCAAAACTATTGCTCCCATAGTATTGCAAC 192
              |||      |||  |||         ||||||  |||      |||      ||||||
Db        123 TyrAsnLeuProCysProArgSerLeuLeuAlaIleLeuSerProProLysCysCysAsn 142

Qy        191 TCTTCAGCGATTCTTTCTGCGTATTTAGCGTAATCATCGCCATGC-----------CTT 144
              |||   ||||||                                ||   |||          |||
Db        143 SerProAlaIleSerCysGlnLeuArgArgCysCysSerGlyCysProSerIleAspLeu 162

Qy        143 TTGTATTCTATGGAGCTGGTCAGTTTTTCATTGTGTCTTTTTTCGCCGTCTTTACCAAAA 84
                |||:::  :::  :::                :::          ||||||
Db        163 AsnSerSerLeuArgIleSerThrLeuGluArgArgValLeuProPheSerLeuTrpVal 182

Qy         83 ACAAGCACCTCAAACAAATCCAATAAATCACTAGATTCCAATTGCTTTAAAAATTCCAAG 24
              :::|||   :::      :::|||   |||                |||::::::   |||:::
Db        183 SerSerArgAlaLysPheAlaAsnCysSer---SerLeuGlnCysTrpArgLysSerArg 201

Qy         23 TCTCTATCATAT 12
              |||   |||:::
Db        202 SerGluSerPhe 205
```

RESULT 12
D85481
hypothetical protein htgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85481
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85481
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-206 <STO>
A;Cross-references: GB:AE005174; NID:g12512690; PIDN:AAG54312.1; GSPDB:GN00145; UWGP:2
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: htgA
C;Superfamily: Escherichia coli heat shock protein htgA

Alignment Scores:

Pred. No.:	0.0104	Length:	206
Score:	127.50	Matches:	63
Percent Similarity:	38.84%	Conservative:	24
Best Local Similarity:	28.12%	Mismatches:	96
Query Match:	9.46%	Indels:	41
DB:	2	Gaps:	7

US-09-732-091-3 (1-759) x D85481 (1-206)

```
Qy        629 ACGCCTGTAATGATCCAGCCAACAGGACCTGTTAAAAAGCTCAGAGTTCTTGTAAGCACC 570
              ||||||          ||||||||||||
Db          5 ThrProPheSerAlaAlaProThrGlyProProSer----------------------- 16

Qy        569 TGATTGCCCGCA---AGCGATAAACCACGCCCTAGAATGGTTTTTGCGACCGCATTCGCA 513
              |||||||         ||||||  |||      :::  |||
Db         17 ------ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaTrpValArg 34

Qy        512 ACAATGACAGCTAATTGATAAGATTTAAAA------CCCCCCATTTTAAACAGCGTTAAA 459
              |||      :::      :::  :::         |||      |||:::
Db         35 LysMetArgValSerTrpLeuGluSerArgCysAspThrProPheAlaAsnAsn------ 52

Qy        458 GTCGCCGCGCTTAAGGCTTGTCTGTTTAAATTGTCCGTGTTTTTTATGGATAATTCATCG 399
                                          ||||||                 :::||||||
Db         53 ----------------------------LeuSerPheIleSerSerGlySerSerSer 62

Qy        398 CACATTTCTTTCACT--------------------------------TCTTCATCATCC 372
              |||||||||                                      |||||||||
Db         63 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuCysSerSerSer 82

Qy        371 ATTTCTTCCAAACTTCTTTCTAAGATTTTAGAAAGCATGTTTTGTTCAATTAAAGTCGTT 312
              |||      ::::::|||   :::         |||      |||||||||      |||
Db         83 IlePhePheGlnValLeuArgArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 102

Qy        311 TCAGTTTTCTTGTTGTAATTGACCTTTAATTTATCGCACACATCGCATAAAATCTCTTTG 252
              :::  |||       :::||||||         ||||||  |||::::::
Db        103 AspIleSerLeuSerAlaPheSerPheAsnArgPheGluThrSerSerLysMetAlaArg 122

Qy        251 TATAAGACTCCTTCGCCTTTAATGAAACTCGCAAAACTATTGCTCCCATAGTATTGCAAC 192
              |||      |||  |||         ||||||  |||      |||      ||||||
Db        123 TyrAsnLeuProCysProArgSerLeuLeuAlaIleLeuSerProProLysCysCysAsn 142

Qy        191 TCTTCAGCGATTCTTTCTGCGTATTTAGCGTAATCATCGCCATGC-----------CTT 144
              |||   ||||||                                ||   |||          |||
Db        143 SerProAlaIleSerCysGlnLeuArgArgCysCysSerGlyCysProSerIleAspLeu 162

Qy        143 TTGTATTCTATGGAGCTGGTCAGTTTTTCATTGTGTCTTTTTTCGCCGTCTTTACCAAAA 84
                |||:::  :::  :::                :::          ||||||
Db        163 AsnSerSerLeuArgIleSerThrLeuGluArgArgValLeuProPheSerLeuTrpVal 182

Qy         83 ACAAGCACCTCAAACAAATCCAATAAATCACTAGATTCCAATTGCTTTAAAAATTCCAAG 24
              :::|||   :::      :::|||   |||                |||::::::   |||:::
Db        183 SerSerArgAlaLysPheAlaAsnCysSer---SerLeuGlnCysTrpArgLysSerArg 201

Qy         23 TCTCTATCATAT 12
              |||   |||:::
Db        202 SerGluSerPhe 205
```

RESULT 13
A40623
heat shock protein htgA - Escherichia coli (strain K-12)
N;Alternate names: heat shock protein Y
C;Species: Escherichia coli
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 01-Mar-2002
C;Accession: A40623; A56688; D64721; S28460

```
R;Missiakas, D.; Georgopoulos, C.; Raina, S.
J. Bacteriol. 175, 2613-2624, 1993
A;Title: The Escherichia coli heat shock gene htpY: mutational analysis, cloning, sequen
A;Reference number: A40623; MUID:93239687; PMID:8478327
A;Accession: A40623
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-196 <MIS>
A;Note: sequence extracted from NCBI backbone (NCBIN:130422, NCBIP:130423)
R;James, R.; Dean, D.O.; Debbage, J.
DNA Seq. 3, 327-332, 1993
A;Title: Five open reading frames upstream of the dnaK gene of Escherichia coli.
A;Reference number: A56688; MUID:94003405; PMID:8400364
A;Accession: A56688
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 36-196 <JAM>
A;Cross-references: EMBL:X67700; NID:g41754; PIDN:CAA47932.1; PID:g41757
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: D64721
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-196 <BLAT>
A;Cross-references: GB:AE000112; GB:U00096; NID:g1786192; PIDN:AAC73123.1; PID:g1786194;
A;Experimental source: strain K-12, substrain MG1655
A;Note: it is uncertain whether Met-1 (CTG) or Met-36 is the initiator
C;Genetics:
A;Gene: htgA; htpY
A;Map position: 0 min
A;Start codon: CTG
C;Superfamily: Escherichia coli heat shock protein htgA
C;Keywords: heat shock; stress-induced protein

Alignment Scores:
Pred. No.:                0.0295          Length:          196
Score:                    122.00          Matches:          51
Percent Similarity:       38.41%          Conservative:     12
Best Local Similarity:    31.10%          Mismatches:       65
Query Match:              9.05%           Indels:           36
DB:                       2               Gaps:              5

US-09-732-091-3 (1-759) x A40623 (1-196)

Qy        629 ACGCCTGTAATGATCCAGCCAACAGGACCTGTTAAAAAGCTCAGAGTTCTTGTAAGCACC 570
              ||||||            ||||||||||||
Db          5 ThrProPheSerAlaAlaProThrGlyProProSer------------------------ 16

Qy        569 TGATTGCCCGCA---AGCGATAAACCACGCCCTAGAATGGTTTTTGCGACCGCATTCGCA 513
              |||||||        ||||||   |||      :::   |||
Db         17 ------ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaTrpValArg 34

Qy        512 ACAATGACAGCTAATTGATAAGATTTAAAA------CCCCCCATTTTAAACAGCGTTAAA 459
              |||      :::    :::   |||       |||     |||:::
Db         35 LysMetArgValSerTrpLeuGluSerLysCysAspThrProPheAlaAsnAsn------ 52

Qy        458 GTCGCCGCGCTTAAGGCTTGTCTGTTTAAATTGTCCGTGTTTTTTATGGATAATTCATCG 399
                                        ||||||               :::||||||
Db         53 ----------------------------LeuSerPheIleSerSerGlySerSerSer 62

Qy        398 CACATTTCTTTCACT--------------------------------TCTTCATCATCC 372
              |||||||||                                    |||||||||
Db         63 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuCysSerSerSer 82

Qy        371 ATTTCTTCCAAACTTCTTTCTAAGATTTTAGAAAGCATGTTTTGTTCAATTAAAGTCGTT 312
              |||      :::::|||   :::        |||      |||||||||      |||
Db         83 IlePhePheGlnValLeuArgArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 102

Qy        311 TCAGTTTTCTTGTTGTAATTGACCTTTAATTTATCGCACACATCGCATAAAATCTCTTTG 252
                 :::   |||         :::||||||         ||||||   |||::::::

Db        103 AspIleSerLeuSerAlaPheSerPheAsnArgPheGluThrSerSerLysMetAlaArg 122

Qy        251 TATAAGACTCCTTCGCCTTTAATGAAACTCGCAAAACTATTGCTCCCATAGTATTGCAAC 192
              |||      |||  |||          ||||||   |||      |||      ||||||
Db        123 TyrAsnLeuProCysProArgSerLeuLeuAlaIleLeuSerProProLysCysCysAsn 142

Qy        191 TCTTCAGCGATT 180
              |||   ||||||
Db        143 SerProAlaIle 146

RESULT 14
C89910
hypothetical protein SA1181 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: C89910
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: C89910
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1009 <KUR>
A;Cross-references: GB:BA000018; PID:g13701144; PIDN:BAB42439.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA1181

Alignment Scores:
Pred. No.:                2.56            Length:          1009
Score:                    98.50           Matches:           66
Percent Similarity:       37.88%          Conservative:      45
Best Local Similarity:    22.53%          Mismatches:        83
Query Match:              7.30%           Indels:            99
DB:                       2               Gaps:              16

US-09-732-091-3 (1-759) x C89910 (1-1009)

Qy         13 TATGATAGAGACTTGGAATTT----------------------TTAAAGCAATTGGAA 48
              |||:::   |||||||||                        :::   :::|||
Db        708 TyrAsnAsnAspLeuGluLeuAsnGluGlnSerIleGluMetGluMetSerArgLeuAsn 727

Qy         49 TCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTTTTTTGGTAAAGACGGCGAAAAAAGA 108
                 :::|||      |||:::  |||::::::  :::      :::   |||:::
Db        728 LeuThrAspAspAsnAspIleAsnGluIleIleAlaTrpArgGlyGluGlnGluGlu--- 746

Qy        109 CACAATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAA 168
                                      :::|||   ||||||      |||   |||   |||
Db        747 ----------------------LeuGluGlnLysArg------AspThrTyrLysLys 756

Qy        169 TACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTC 228
                                    |||:::                        |||
Db        757 ArgTyrHisGluPheGluMetGluIleAla--------------ArgLeuGluSerLeu 771

Qy        229 ATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAG 288
              |||                  ||||||:::|||             ||||||||||||
Db        772 ThrLysAsp--------------LysGluLeuLeuAspSer------AspLysLeuLys 784

Qy        289 GTCAATTACAACAAGAAAACTGAA------ACGACTTTAATTGAA--------------- 327
                 |||  :::|||   |||       |||||||||:::
Db        785 AspGluTyrGluGlnLysLysGluLysMetAsnThrLeuIleAspGluTyrSerAlaVal 804

Qy        328 --------------------------------CAAAACATGCTTTCTAAAATCTTAGAA 354
                                               |||::::::::::|||   |||
Db        805 HisTyrGlnCysGlnAsnAsnIleAsnLysThrGlnSerIleValSerHisIleAsnTyr 824

Qy        355 AGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATA 414
                 :::   |||       |||   :::|||:::   ::::::   :::   :::|||
Db        825 LeuAsnGlnGluLeuLysAspGlnGlnGluIlePheGlnLeuAlaGluIleValSerGly 844
```

```
Qy        415 AAAAACACGGACAATTTAAAC----------------------------AGACAAGCC 444
Db        845 LysAsnAsnLysAsnLeuThrLeuGluAsnPheValLeuIleTyrTyrLeuAspGlnIle 864

Qy        445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCT 504
Db        865 IleAlaGlnAlaAsnLeuArgLeuAlaThrMetSerAsp---AsnArgTyrGlnLeu--- 882

Qy        505 GTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCG--------- 555
Db        883 ------------------IleArgArgGluAlaValSerHisGlyLeuSerGlyLeuGlu 896

Qy        556 ---------------CTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
Db        897 IleAspValPheAspLeuHisSerAsnLys------SerArgHisIleSerSerLeuSer 914

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
Db        915 Gly----------------------------------------------GlyGluThr 918

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
Db        919 PheGlnSerSerLeuAlaLeuAlaLeuGlyLeuSerGluIle---------ValGlnGln 935

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
Db        936 GlnSerGly---------GlyIleSerLeuGluSerIle 945
```

RESULT 15
AE0360
hypothetical protein YPO2963 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AE0360
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AE0360
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-282 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92208.1; PID:g15980920; GSPDB:GN00175
C;Genetics:
A;Gene: YPO2963

Alignment Scores:

Pred. No.:	2.8	Length:	282
Score:	98.00	Matches:	58
Percent Similarity:	39.15%	Conservative:	34
Best Local Similarity:	24.68%	Mismatches:	87
Query Match:	7.26%	Indels:	56
DB:	2	Gaps:	13

US-09-732-091-3 (1-759) x AE0360 (1-282)

```
Qy         55 GATTTATTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAA---AGACAC 111
Db          7 GluValLeuAspAlaValGluArgAsnProTyrSerAsnSerGluAspLysThrArgGln 26

Qy        112 AATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGAT---GATTACGCTAAA 168
Db         27 ArgAsnAsnValMetArgSerIleMetTyrAsnLysGlnAsnGluLeuAspGlnValLys 46

Qy        169 TACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGT--- 225
Db         47 ----------------------------------------AsnSerIleAlaSerLys 52

Qy        226 TTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTA 285
Db         53 PheIleLysTrpSer------LeuIleLysAspGlnPheIleSerGluValThrArgTyr 70

Qy        286 AAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAA 345
Db         71 SerSerAspIleThrAlaLysValAsnThrGlyGluIleThrThrGluThrAlaIleLys 90

Qy        346 ATCTTAGAAAGAAGTTTG--------------------GAAGAAATGGATGATGAAGAA 384
Db         91 LeuLeuAspLysGluLeuSerAspLeuArgAsnGlnAspGluGluMetThrArgGluArg 110

Qy        385 GTGAAAGAA---ATGTGCGATGAATTATCCATAAAAAACACGGACAATTTAAACAGACAA 441
Db        111 ValLysGlnAlaValIleIleLysAlaSerValLysGluAsnAsnAsnValGluArgAsn 130

Qy        442 GCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG---GGTTTTAAATCTTATCAA 498
Db        131 GluVal------------IleAsnLeuValPheAlaGlyValGlyPheValThrAlaGly 146

Qy        499 TTAGCTGTCATT------------GTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGT 546
Db        147 LeuGlnIleValAlaGlyValGlyMetValGlySerValValGlySerIleProGlyThr 166

Qy        547 GGTTTATCGCTTGCGGGC---AATCAGGTGCTTACAAGAACTCTGAGCTTTTTA------ 597
Db        167 LeuLeuIleLeuHisGlyAlaAsnAsnIleValGluSerGlyTyrTyrIleLeuTyrArg 186

Qy        598 ---------ACAGGTCCTGTTGGCTGGATCATTACAGGCGTA------------------ 630
Db        187 GluSerTyrThrGlyProValLysPheValTyrGluGlyValGlyAlaGlnPheGlyLeu 206

Qy        631 ------------------------TGGACAGCGATTGATATTGCA 651
Db        207 SerLysSerAspSerAlaIleMetTyrThrAlaValAspValAla 221
```

Search completed: July 5, 2004, 03:08:17
Job time : 27.5 secs

GenCore version 5.1.6

CM nucleic - protein search, using frame_plus_n2p model

Run on: July 5, 2004, 02:52:17 ; Search time 14.5 Seconds
(without alignments)
5451.207 Million cell updates/sec

Title: US-09-732-091-3
Perfect score: 1349
Sequence: 1 atggcatacaaatatgatag.....cgttgcaaatagaatccatt 759

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09732091/runat_02072004_182406_19007/app_query.fasta_1.903
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09732091_@CGN_1_1_12_@runat_02072004_182406_19007 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1279	94.8	253	1	YF88_HELPY	O26107 helicobacte
	2	1270	94.1	253	1	YF88_HELPJ	Q9zj24 helicobacte
	3	637	47.2	209	1	YF87_HELPY	O26106 helicobacte
	4	630	46.7	209	1	YF87_HELPJ	Q9zj25 helicobacte
	5	265	19.6	237	1	YAAW_ECOLI	P75617 escherichia
	6	262	19.4	237	1	YAAW_ECO57	P58316 escherichia
c	7	122	9.1	196	1	HTGA_ECOLI	P28697 escherichia
	8	105	7.8	726	1	HS9A_BRARE	Q90474 brachydanio
	9	97	7.2	916	1	SECA_TREPA	O83394 treponema p
	10	96.5	7.2	373	1	BIOF_AQUAE	O66875 aquifex aeo
	11	95.5	7.1	344	1	ABIC_LACLA	Q01457 lactococcus
	12	95	7.0	517	1	YC09_CAMJE	Q9pn86 campylobact
	13	95	7.0	3214	1	BPA1_HUMAN	Q03001 homo sapien
	14	94.5	7.0	1433	1	REST_CHICK	O42184 gallus gall
	15	93.5	6.9	802	1	PAC_BACME	Q60136 bacillus me
	16	92.5	6.9	802	1	PAC_ARTVI	P31956 arthrobacte
	17	92.5	6.9	892	1	RCQ1_CAEEL	P46064 caenorhabdi
	18	91.5	6.8	399	1	RPSD_THEMA	P77994 thermotoga
	19	91.5	6.8	472	1	GATB_METJA	Q57624 methanococc
	20	91.5	6.8	719	1	HS9A_HORSE	Q9gkx7 equus cabal
	21	91.5	6.8	1769	1	YJK9_YEAST	P42945 saccharomyc
	22	91	6.7	284	1	TPM1_CIOIN	Q07068 ciona intes
	23	90	6.7	404	1	LA_BOVIN	P10881 bos taurus
	24	89.5	6.6	280	1	Y705_CHLPN	Q9z7k0 chlamydia p
	25	89.5	6.6	633	1	GATE_SULSO	Q97zh6 sulfolobus
	26	89.5	6.6	869	1	AMPN_HAEIN	P45274 haemophilus
	27	89	6.6	386	1	NESG_HUMAN	Q9ul16 homo sapien
	28	88.5	6.6	732	1	HS9A_CRIGR	P46633 cricetulus
	29	88	6.5	371	1	RPSD_STRMU	O33662 streptococc
	30	88	6.5	670	1	SSM4_SCHPO	O42667 schizosacch
	31	87.5	6.5	728	1	HS9A_CHICK	P11501 gallus gall
	32	87.5	6.5	731	1	HS9A_HUMAN	P07900 homo sapien
	33	87.5	6.5	732	1	HS9A_MOUSE	P07901 mus musculu
	34	87.5	6.5	732	1	HS9A_PIG	O02705 sus scrofa
	35	87	6.4	607	1	HTPG_FUSNN	Q8rgh4 fusobacteri
	36	86.5	6.4	619	1	PRIA_HELPJ	Q9zke4 helicobacte
	37	86.5	6.4	686	1	HMCT_HELPJ	Q9zl53 helicobacte
	38	86.5	6.4	1076	1	CARB_ARCFU	O28994 archaeoglob
	39	86.5	6.4	1102	1	CARB_STRCO	Q9kxr6 streptomyce
c	40	86	6.4	313	1	O5DI_HUMAN	Q8ngl1 homo sapien
	41	86	6.4	619	1	PRIA_HELPY	O25149 helicobacte
	42	86	6.4	633	1	YN40_ARCFU	O27944 archaeoglob
	43	86	6.4	1164	1	KEL1_YEAST	P38853 saccharomyc
	44	86	6.4	1169	1	EX5B_BORBU	O51578 borrelia bu
	45	85.5	6.3	411	1	HFLK_BUCAP	Q8k914 buchnera ap

ALIGNMENTS

RESULT 1
YF88_HELPY
ID YF88_HELPY STANDARD; PRT; 253 AA.
AC O26107;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein HP1588.
GN HP1588.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
CC -!- SIMILARITY: Belongs to the UPF0174 family.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; AE000656; AAD08627.1; -.
DR PIR; D64718; D64718.
DR TIGR; HP1588; -.

```
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   253 AA;  28417 MW;  00E15A38B1A2036A CRC64;

Alignment Scores:
Pred. No.:              7.56e-96        Length:         253
Score:                  1279.00         Matches:        253
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            94.81%          Indels:         0
DB:                     1               Gaps:           0

US-09-732-091-3 (1-759) x YF88_HELPY (1-253)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              |||||||||||||||||||||||||||||||||||||||
Db        241 AlaAsnGlyAspLysLysSerLeuGlnIleGluSerIle 253

RESULT 2
YF88_HELPJ
ID   YF88_HELPJ     STANDARD;      PRT;   253 AA.
AC   Q9ZJ24;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein JHP1494.
GN   JHP1494.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- SIMILARITY: Belongs to the UPF0174 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001571; AAD07073.1; -.
DR   PIR; B71800; B71800.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   253 AA;  28475 MW;  127158B2B1A2D36A CRC64;

Alignment Scores:
Pred. No.:              4.03e-95        Length:         253
Score:                  1270.00         Matches:        251
Percent Similarity:     99.60%          Conservative:   1
Best Local Similarity:  99.21%          Mismatches:     1
Query Match:            94.14%          Indels:         0
DB:                     1               Gaps:           0

US-09-732-091-3 (1-759) x YF88_HELPJ (1-253)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140
```

```
Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              |||||||   ||||||||||||||||||||||||:::
Db        241 AlaAsnGluAspLysLysSerLeuGlnIleGluSerVal 253
```

```
RESULT 3
YF87_HELPY
ID   YF87_HELPY     STANDARD;      PRT;   209 AA.
AC   O26106;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein HP1587.
GN   HP1587.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
RN   [2]
RP   CONCEPTUAL TRANSLATION.
RA   Bairoch A.;
RL   Unpublished observations (OCT-2001).
CC   -!- SIMILARITY: Belongs to the UPF0174 family.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC       frameshift in position 160.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000656; AAD08626.1; ALT_FRAME.
DR   TIGR; HP1587; -.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   209 AA;  23069 MW;  F98D3FB8F3F62323 CRC64;
```

```
Alignment Scores:
Pred. No.:              5.46e-44        Length:         209
Score:                  637.00          Matches:        123
Percent Similarity:     76.47%          Conservative:   33
Best Local Similarity:  60.29%          Mismatches:     44
Query Match:            47.22%          Indels:         4
DB:                     1               Gaps:           2

US-09-732-091-3 (1-759) x YF87_HELPY (1-209)

Qy        112 AATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATAC 171
              ||||||   ||||||:::|||   ||||||||||||:::|||   ||||||||||||||||
Db          2 AsnGluAspLeuThrAsnSerThrGluTyrLysArgTyrGlyHisAspTyrAlaLysTyr 21

Qy        172 GCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATT 231
              ||||||||||||||||||:::||||||   ||||||||||||:::|||
Db         22 ProArgArgIleAlaGluGluLeuGlnHisTyrGlyGlyAsnSerPheAlaAsnPhePhe 41

Qy        232 AAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTC 291
              :::   ||||||||||||||||||||||||||||||   ||||||   ||||||||
Db         42 ArgAspGluGlyValLeuTyrLysGluIleLeuCysAspAlaCysAspHisLeuLysVal 61

Qy        292 AATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTA 351
              |||||||||:::::::::   |||:::||||||||||||||||||||||||:::|||
Db         62 AsnTyrAsnGluGluSerAlaThrSerLeuIleGluGlnAsnMetLeuSerLysLeuLeu 81

Qy        352 GAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCC 411
              :::   ||||||||:::|||          |||:::||||||:::|||:::||||||
Db         82 LysAspSerLeuGluLysMetSerArgArgGluIleLysGluLeuCysAsnGluLeuGly 101

Qy        412 ATAAAAAACACGGACAATTTA---------AACAGACAAGCCTTAAGCGCGGCGACTTTA 462
              :::   |||   ||  :::          |||:::|||   |||   |||:::||||||
Db        102 MetThrAsnIleAspLysValIleGlyGluAsnLysGlnValLeuIleAlaSerThrLeu 121

Qy        463 ACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCG 522
              ||||||||||||   ||||||      ||||||   |||||||||   |||||:::|||
Db        122 ThrLeuPheLysAlaGlyGlySerHisSerTyrAlaLeuAlaValSerValAlaAspAla 141

Qy        523 GTCGCAAAAACCATTCTAGGGCGT---GGTTTATCGCTTGCGGGCAATCAGGTGCTTACA 579
              :::   :::      ||||||                :::   |||         |||
Db        142 MetValArgGlnThrLeuGlyHis***AlaCysTyrValValGlyLysValAlaLeuLys 161

Qy        580 AGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCG 639
              :::||||||      |||   ||||||:::|||||:::|||||||||         :::
Db        162 LysThrLeuGlyValLeuAlaGlyProIleGlyTrpValIleThrGlyAlaLeuValSer 181

Qy        640 ATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACT 699
              |||::::::||||||||||||||||||||||||:::|||||||||:::::::::||||||
Db        182 IleAsnLeuAlaGlyProAlaTyrArgValThrValProAlaCysValLeuIleAlaThr 201

Qy        700 TTACGCCTAAAA 711
              ||||||||||||
Db        202 LeuArgLeuLys 205
```

```
RESULT 4
YF87_HELPJ
ID   YF87_HELPJ     STANDARD;      PRT;   209 AA.
AC   Q9ZJ25;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein JHP1493.
GN   JHP1493.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
```

```
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- SIMILARITY: Belongs to the UPF0174 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001571; AAD07072.1; -.
DR   PIR; A71800; A71800.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   209 AA;  22907 MW;  C6EC950CDD424CAF CRC64;

Alignment Scores:
Pred. No.:                2.01e-43        Length:          209
Score:                    630.00          Matches:         122
Percent Similarity:       75.49%          Conservative:    32
Best Local Similarity:    59.80%          Mismatches:      46
Query Match:              46.70%          Indels:          4
DB:                       1               Gaps:            2

US-09-732-091-3 (1-759) x YF87_HELPJ (1-209)

Qy        112 AATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATAC 171
              ||||||:::|||||||||      ||||||:::|||:::|||  ||||||||||||||||
Db          2 AsnGluGluLeuThrSerLeuThrGluTyrGlnArgTyrGlyHisAspTyrAlaLysTyr 21

Qy        172 GCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATT 231
              ||||||||||||||||||||   ||||||   |||||||||||:::|||
Db         22 ProArgArgIleAlaGluGluLeuGlnArgTyrGlyGlyAsnSerPheAlaAsnPhePhe 41

Qy        232 AAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTC 291
              :::  ||||||||||||||||||||||||||||||||||   ||||||   |||   :::
Db         42 ArgAspGluGlyValLeuTyrLysGluIleLeuCysAspAlaCysAspHisLeuAspIle 61

Qy        292 AATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTA 351
              |||||||||:::::::::   |||:::|||||||||||||||||||||||||:::|||
Db         62 AsnTyrAsnGluArgSerAlaThrSerLeuIleGluGlnAsnMetLeuSerLysLeuLeu 81

Qy        352 GAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCC 411
              :::   |||||||||:::|||        |||:::|||||:::||||||   |||
Db         82 LysAspSerLeuGluLysMetSerGlyArgGluIleLysGluLeuCysAspGlyLeuGly 101

Qy        412 ATAAAAAACACGGACAATTTA---------AACAGACAAGCCTTAAGCGCGGCGACTTTA 462
              :::   |||   |||   :::         |||:::|||   |||   |||:::   |||
Db        102 MetProAsnIleAspLysValIleGlyGluAsnLysGlnValLeuIleAlaSerValLeu 121

Qy        463 ACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCG 522
              ||||||||||||   ||||||      ||||||   |||||||||   |'|||||:::|||
Db        122 ThrLeuPheLysAlaGlyGlySerHisSerTyrAlaLeuAlaValAlaValAlaAspAla 141

Qy        523 GTCGCAAAAACCATTCTAGGGCGTGGTTTA---TCGCTTGCGGGCAATCAGGTGCTTACA 579
              :::   :::      |||||   ||||||   |||:::   |||      |||
Db        142 MetValArgGlnThrLeuGlyHisGlyLeuSerSerValValGlyLysValAlaLeuLys 161

Qy        580 AGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCG 639
              :::||||||      |||   ||||||:::||||||:::|||||||||      :::
Db        162 LysThrLeuAspIleLeuAlaGlyProIleGlyTrpValIleThrGlyAlaLeuValSer 181

Qy        640 ATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACT 699
              |||::::::|||||||||||||||||||||||:::|||||||||::::::|||||||||
```

```
Db        182 IleAsnLeuAlaGlyProAlaTyrArgValThrValProAlaCysValLeuValAlaThr 201

Qy        700 TTACGCCTAAAA 711
              ||||||   |||
Db        202 LeuArgLysLys 205

RESULT 5
YAAW_ECOLI
ID   YAAW_ECOLI     STANDARD;      PRT;   237 AA.
AC   P75617; Q47290;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein yaaW.
GN   YAAW OR B0011.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94003405; PubMed=8400364;
RA   James R., Dean D.O., Debbage J.;
RT   "Five open reading frames upstream of the dnaK gene of E. coli.";
RL   DNA Seq. 3:327-332(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- SIMILARITY: Belongs to the UPF0174 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67700; CAA47934.1; -.
DR   EMBL; AE000112; AAC73122.1; -.
DR   PIR; C64721; C64721.
DR   EcoGene; EG14340; yaaW.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT    190    190       L -> F (IN REF. 1).
SQ   SEQUENCE   237 AA;  26665 MW;  A36682B2EA116747 CRC64;

Alignment Scores:
Pred. No.:                6.16e-14        Length:          237
Score:                    265.00          Matches:         77
Percent Similarity:       48.99%          Conservative:    44
Best Local Similarity:    31.17%          Mismatches:      90
Query Match:              19.64%          Indels:          36
DB:                       1               Gaps:            6

US-09-732-091-3 (1-759) x YAAW_ECOLI (1-237)

Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
              |||      |||   ||||||:::||||||:::                  |||   :::
Db          5 TyrLeuAsnAspSerAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeuAlaAsn 24

Qy         67 TTGTTTGAGGTGCTTGTTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
                  :::|||         :::   |||:::   |||   :::       |||
Db         25 PheAlaArgLeuLeuThrHisAsnGluLysGlyLysThrArgLeuSerSerValLeuMet 44
```

```
Qy        127 AGCTCCATAGAATACAAAAGG---------CATGGCGATGATTACGCTAAATACGCAGAA 177
Db         45 ArgAsnGluLeuPheLysSerMetGluGlyHisProGluGlnHisArgArgAsnTrpGln 64

Qy        178 AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGC 237
Db         65 LeuIleAlaGlyGluLeuGlnHisPheGlyGlyAspSerIleAlaAsnLysLeuArgGly 84

Qy        238 GAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTAC 297
Db         85 HisGlyLysLeuTyrArgAlaIleLeuLeuAspValSerLysArgLeuLysLeuLysAla 104

Qy        298 AACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGA 357
Db        105 AspLysGluMetSerThrPheGluIleGluGlnGlnLeuLeuGluGlnPheLeuArgAsn 124

Qy        358 AGTTTGGAAGAAATGGATGATGAA--------------------------------GAA 384
Db        125 ThrTrpLysLysMetAspGluGluHisLysGlnGluPheLeuHisAlaValAspAlaArg 144

Qy        385 GTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCC 444
Db        145 ValAsnGluLeuGluGluLeuLeuProLeuLeuMetLysAspLysLeuLeuAlaLysGly 164

Qy        445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCT 504
Db        165 ValSer------------------------------HisLeuLeuSerSerGlnLeuThr 174

Qy        505 GTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGC 564
Db        175 ArgIleLeuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu---LeuArgGly 193

Qy        565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACA 624
Db        194 ------------------------AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205

Qy        625 GGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGC 684
Db        206 GlyValLysAla---------ValSerGlySerAlaTyrArgValThrIleProAlaVal 222

Qy        685 ATTGTGGTTGCCACTTTACGC 705
Db        223 LeuGlnIleAlaCysLeuArg 229
```

```
RESULT 6
YAAW_ECO57
ID   YAAW_ECO57     STANDARD;      PRT;   237 AA.
AC   P58316;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein yaaW.
GN   YAAW OR Z0011 OR ECS0012.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- SIMILARITY: Belongs to the UPF0174 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE005178; AAG54311.1; -.
DR   EMBL; AP002550; BAB33434.1; -.
DR   PIR; C85481; C85481.
DR   PIR; C90630; C90630.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   237 AA;  26681 MW;  A25482B2EA116759 CRC64;
```

```
Alignment Scores:
Pred. No.:                1.08e-13      Length:          237
Score:                    262.00        Matches:         76
Percent Similarity:       48.99%        Conservative:    45
Best Local Similarity:    30.77%        Mismatches:      90
Query Match:              19.42%        Indels:          36
DB:                       1             Gaps:            6

US-09-732-091-3 (1-759) x YAAW_ECO57 (1-237)

Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
Db          5 TyrLeuAsnAspSerAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeuAlaAsn 24

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
Db         25 PheAlaArgLeuLeuThrHisAsnGluLysGlyLysThrArgLeuSerSerValLeuMet 44

Qy        127 AGCTCCATAGAATACAAAAGG---------CATGGCGATGATTACGCTAAATACGCAGAA 177
Db         45 ArgAsnGluLeuPheLysSerMetGluGlyHisProGluGlnHisArgArgAsnTrpGln 64

Qy        178 AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGC 237
Db         65 LeuIleAlaGlyGluLeuGlnHisPheGlyGlyAspSerIleAlaAsnLysLeuArgGly 84

Qy        238 GAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTAC 297
Db         85 HisGlyLysLeuTyrArgAlaIleLeuLeuAspValSerLysArgLeuLysLeuLysAla 104

Qy        298 AACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGA 357
Db        105 AspLysGluMetSerThrPheGluIleGluGlnGlnLeuLeuGluGlnPheLeuArgAsn 124

Qy        358 AGTTTGGAAGAAATGGATGATGAA--------------------------------GAA 384
Db        125 ThrTrpLysLysMetAspGluGluHisLysGlnGluPheLeuHisAlaValAspAlaArg 144

Qy        385 GTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCC 444
Db        145 ValAsnGluLeuGluGluLeuLeuProLeuLeuMetLysAspLysLeuLeuAlaLysGly 164

Qy        445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCT 504
Db        165 ValSer------------------------------HisLeuLeuSerSerGlnLeuThr 174

Qy        505 GTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGC 564
```

```
Db        175 ArgIleLeuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu---LeuArgGly 193

Qy        565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACA 624

Db        194 ----------------------AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205

Qy        625 GGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGC 684

Db        206 GlyValLysAla---------ValSerGlySerSerTyrArgValThrIleProAlaVal 222

Qy        685 ATTGTGGTTGCCACTTTACGC 705

Db        223 LeuGlnIleAlaCysLeuArg 229

RESULT 7
HTGA_ECOLI
ID   HTGA_ECOLI     STANDARD;      PRT;   196 AA.
AC   P28697;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Very hypothetical heat shock protein htgA (Heat shock protein htpY).
GN   HTGA OR HTPY OR B0012.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94003405; PubMed=8400364;
RA   James R., Dean D.O., Debbage J.;
RT   "Five open reading frames upstream of the dnaK gene of E. coli.";
RL   DNA Seq. 3:327-332(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RX   MEDLINE=93239687; PubMed=8478327;
RA   Missiakas D., Georgopoulos C., Raina S.;
RT   "The Escherichia coli heat shock gene htpY: mutational analysis,
RT   cloning, sequencing, and transcriptional regulation.";
RL   J. Bacteriol. 175:2613-2624(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92334977; PubMed=1630901;
RA   Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA   Isono K., Mizobuchi K., Nakata A.;
RT   "Systematic sequencing of the Escherichia coli genome: analysis of
RT   the 0-2.4 min region.";
RL   Nucleic Acids Res. 20:3305-3308(1992).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: NOT KNOWN, REQUIRED FOR HIGH-TEMPERATURE GROWTH OF
CC       E.COLI. IT IS POSSIBLE THAT HTGA PROTEIN SOMEHOW REGULATES EITHER
CC       TRANSCRIPTION OF THE RPOH GENE OR THE ACTIVITY OF ITS GENE PRODUCT
CC       SIGMA-32. PLAYS A ROLE OPPOSITE THAT OF DNAK, DNAJ, AND GRPE IN
CC       TERMS OF HEAT SHOCK REGULATION, ANTAGONIZING THE NEGATIVE OF
CC       THESE AND THUS FINE-TUNING THE HEAT SHOCK RESPONSE.
CC   -!- INDUCTION: By high temperatures.
CC   -!- CAUTION: It is uncertain whether Met-1 or Met-36 is the initiator.
CC   -!- CAUTION: Ref.3 sequence differs from that shown due to a
CC       frameshift.
CC   -!- CAUTION: This sequence, according to the EcoSeq database (K. Rudd)
CC       is probably not a real protein; therefore this entry will probably
CC       be deleted in future releases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67700; CAA47932.1; ALT_INIT.
DR   EMBL; L03720; AAA23992.1; -.
DR   EMBL; L03720; AAA23993.1; ALT_INIT.
DR   EMBL; D10483; -; NOT_ANNOTATED_CDS.
DR   EMBL; AE000112; AAC73123.1; -.
DR   PIR; A40623; A40623.
KW   Hypothetical protein; Heat shock.
FT   DOMAIN       57      65       POLY-SER.
SQ   SEQUENCE   196 AA;  21225 MW;  ECA6154160A40993 CRC64;

Alignment Scores:
Pred. No.:              0.0218        Length:          196
Score:                  122.00        Matches:         51
Percent Similarity:     38.41%        Conservative:    12
Best Local Similarity:  31.10%        Mismatches:      65
Query Match:            9.05%         Indels:          36
DB:                     1             Gaps:            5

US-09-732-091-3 (1-759) x HTGA_ECOLI (1-196)

Qy        629 ACGCCTGTAATGATCCAGCCAACAGGACCTGTTAAAAAGCTCAGAGTTCTTGTAAGCACC 570

Db          5 ThrProPheSerAlaAlaProThrGlyProProSer------------------------ 16

Qy        569 TGATTGCCCGCA---AGCGATAAACCACGCCCTAGAATGGTTTTTGCGACCGCATTCGCA 513

Db         17 ------ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaTrpValArg 34

Qy        512 ACAATGACAGCTAATTGATAAGATTTAAAA------CCCCCCATTTTAAACAGCGTTAAA 459

Db         35 LysMetArgValSerTrpLeuGluSerLysCysAspThrProPheAlaAsnAsn------ 52

Qy        458 GTCGCCGCGCTTAAGGCTTGTCTGTTTAAATTGTCCGTGTTTTTTATGGATAATTCATCG 399

Db         53 ----------------------------LeuSerPheIleSerSerGlySerSerSer 62

Qy        398 CACATTTCTTTCACT--------------------------------TCTTCATCATCC 372

Db         63 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuCysSerSerSer 82

Qy        371 ATTTCTTCCAAACTTCTTTCTAAGATTTTAGAAAGCATGTTTTGTTCAATTAAAGTCGTT 312

Db         83 IlePhePheGlnValLeuArgArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 102

Qy        311 TCAGTTTTCTTGTTGTAATTGACCTTTAATTTATCGCACACATCGCATAAAATCTCTTTG 252

Db        103 AspIleSerLeuSerAlaPheSerPheAsnArgPheGluThrSerSerLysMetAlaArg 122

Qy        251 TATAAGACTCCTTCGCCTTTAATGAAACTCGCAAAACTATTGCTCCCATAGTATTGCAAC 192

Db        123 TyrAsnLeuProCysProArgSerLeuLeuAlaIleLeuSerProProLysCysCysAsn 142

Qy        191 TCTTCAGCGATT 180

Db        143 SerProAlaIle 146

RESULT 8
HS9A_BRARE
ID   HS9A_BRARE     STANDARD;      PRT;   726 AA.
AC   Q90474;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE   Heat shock protein HSP 90-alpha.
GN   HSP90A OR HSP90.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lele Z., Hadfi S., Sass J.B., Krone P.H.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 36-136 FROM N.A.
RX   MEDLINE=95071389; PubMed=7980538;
RA   Krone P.H., Sass J.B.;
RT   "HSP 90 alpha and HSP 90 beta genes are present in the zebrafish and
RT   are differentially regulated in developing embryos.";
RL   Biochem. Biophys. Res. Commun. 204:746-752(1994).
CC   -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the heat shock protein 90 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF068773; AAC21567.1; -.
DR   EMBL; L35586; AAA97518.1; -.
DR   PIR; JC2343; JC2343.
DR   HSSP; P07900; 1BYQ.
DR   ZFIN; ZDB-GENE-990415-94; hsp90a.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR001404; Hsp90.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00183; HSP90; 1.
DR   PRINTS; PR00775; HEATSHOCK90.
DR   SMART; SM00387; HATPase_c; 1.
DR   PROSITE; PS00298; HSP90; 1.
KW   Chaperone; ATP-binding; Heat shock.
SQ   SEQUENCE   726 AA;  83561 MW;  F7DEB8EF1FBC9CEC CRC64;

Alignment Scores:
Pred. No.:              0.537          Length:         726
Score:                  105.00         Matches:        61
Percent Similarity:     35.79%         Conservative:   41
Best Local Similarity:  21.40%         Mismatches:     75
Query Match:            7.78%          Indels:         108
DB:                     1              Gaps:           13

US-09-732-091-3 (1-759) x HS9A_BRARE (1-726)

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||      |||        ||||||      |||::::::   ||:::
Db        415 LeuAspLeuPheThrGluLeuAlaGluAspLysAspAsnTyrLysLysTyrTyrGluGln 434

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
                 :::   :::|||:::          |||   |||      ::::::      ::::::
Db        435 PheSerLysAsnIleLysLeuGlyIleHisGluAsp------SerGlnAsnArgLysLys 452

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGT---------------------- 216
              :::::::::   |||:::||||||   |||   |||
Db        453 LeuSerAspLeuLeuArgTyrTyrThrSerAlaSerGlyAspGluMetValSerLeuLys 472

Qy        217 ---TTTGCGAGT----------------------------TTCATTAAAGGCGAA--- 240
                 :::   |||                          :::|||   ||||||
Db        473 AspTyrValSerArgMetLysAspThrGlnLysHisIleTyrTyrIleThrGlyGluThr 492

Qy        241 ------------------------------------------------------GGAGTC 246

                                                                      |||
Db        493 LysAspGlnValAlaAsnSerAlaPheValGluArgLeuArgLysAlaGlyLeuGluVal 512

Qy        247 TTATACAAA---GAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAG 303
              :::|||      |||   :::   :::   |||   :::            |||:::
Db        513 IleTyrMetIleGluProIleAspGluTyrCysValGlnGlnLeuLysGluTyrAspGly 532

Qy        304 AAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTG 363
              |||                        |||::::::|||   :::   :::      |||
Db        533 Lys-----------------------AsnLeuValSer---ValThrLysGluGlyLeu 543

Qy        364 GAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACG 423
              |||      :::|||||||||   |||:::      |||||||||      |||
Db        544 GluLeuProGluAspGluGluGluLysLysLysGlnAspGluLeuLysAlaLys---Tyr 562

Qy        424 GACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGT 483
              :::||||||   :::
Db        563 GluAsnLeuCysLys------------------------------------------- 567

Qy        484 TTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGG 543
                                                     :::   |||   ||||||
Db        568 ------------------------------------IleMetLysAspIleLeuAsp 574

Qy        544 CGTGGTTTATCG------CTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTA 597
              :::   :::              |||:::::::::::              :::
Db        575 LysLysIleGluLysValThrValSerAsnArgLeuValSerSerProCysCysIleVal 594

Qy        598 ACAGGTCCTGTTGGCTGGATCATTACA------------------------------- 624
              |||        ||||||
Db        595 ThrSerThrTyrGlyTrpThrAlaAsnMetGluArgIleMetLysSerGlnAlaLeuArg 614

Qy        625 ---------------GGCGTATGGACAGCG---------ATTGATATTGCAGGGCCGGCT 660
                            |||      ||||||         ::::::|||
Db        615 AspAsnSerThrMetGlyTyrMetThrAlaLysLysHisLeuGluIleAsn--------- 631

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
                          ||||||      :::|||   ||||||||   |||   :::
Db        632 ---------------ProAlaHisProIleValGluThrLeuArgGluLysAlaGluAla 646

Qy        721 GCCAATGGAGATAAG 735
                     |||||
Db        647 GluLysAsnAspLys 651

RESULT 9
SECA_TREPA
ID   SECA_TREPA     STANDARD;      PRT;   916 AA.
AC   O83394;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Preprotein translocase secA subunit.
GN   SECA OR TP0379.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Nichols;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- FUNCTION: Involved in protein export. Interacts with the secY/secE
CC       subunits. SecA has a central role in coupling the hydrolysis of
CC       ATP to the transfer of pre-secretory periplasmic and outer
CC       membrane proteins across the membrane (By similarity).
```

```
CC   -!- SUBUNIT: Part of the prokaryotic protein translocation apparatus
CC       which comprise secA, secB, secD, secE, secF, secG and secY (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic side of plasma membrane (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the secA family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001217; AAC65365.1; -.
DR   PIR; E71330; E71330.
DR   TIGR; TP0379; -.
DR   InterPro; IPR004027; SEC_C_motif.
DR   InterPro; IPR000185; SecA.
DR   Pfam; PF02810; SEC-C; 1.
DR   Pfam; PF01043; SecA_protein; 1.
DR   PRINTS; PR00906; SECA.
DR   TIGRFAMs; TIGR00963; secA; 1.
DR   PROSITE; PS01312; SECA; 1.
KW   Protein transport; ATP-binding; Membrane; Translocation; Transport;
KW   Complete proteome.
FT   NP_BIND     103    110       ATP (POTENTIAL).
SQ   SEQUENCE   916 AA;  103779 MW;  EA5561F6EE7C65AE CRC64;

Alignment Scores:
Pred. No.:              2.39          Length:          916
Score:                  97.00         Matches:         52
Percent Similarity:     40.17%        Conservative:    44
Best Local Similarity:  21.76%        Mismatches:      67
Query Match:            7.19%         Indels:          76
DB:                     1             Gaps:            13

US-09-732-091-3 (1-759) x SECA_TREPA (1-916)

Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
              :::|||||          :::   :::      |||          :::   ||||||
Db        311 PheLysTyr---------IleHisTyrPheThrGlnAlaLeuArgAlaHisLeuLeuTyr 327

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAA---------------------- 102
                 :::|||              |||||||||:::
Db        328 ArgAlaAspValAspTyrValValLysAspGlyGlnValGlnIleValAspGluPheThr 347

Qy        103 ------------------AAAAGACACAATGAAAAACTGACCAGCTCCATAGAATACAAA 144
                               :::|||:::::::::   |||      :::||||||   |||
Db        348 GlyArgIleLeuGluGlyArgArgTyrSerAspGlyLeuHisGlnAlaIleGluAlaLys 367

Qy        145 AGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTAT 204
                |||                          ||||||||:::     :::
Db        368 GluHis------------------------IleArgIleAlaGlnArgAsnArgThrMet 379

Qy        205 GGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATT--- 261
                 :::   :::|||   :::|||   :::            :::||||||::::::
Db        380 AlaThrIleThrPheGlnAsnPhePheArg-----------MetTyrLysLysLeuSer 395

Qy        262 ---TTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAG------AAAACTGAA 312
                  :::          |||          :::      ||||||      |||   |||
Db        396 GlyMetThrGlyThrAlaAspThrGluAlaLeuGluLeuAsnLysIleTyrLysLeuGlu 415

Qy        313 ACGACTTTAATTGAACAAAACATG---CTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAA 369
                  ::::::      |||:::   ::::::::::::   |||      :::
Db        416 ValValValLeuProThrAsnLeuProValAlaArgValAspGluHisAspValValTyr 435

Qy        370 ATGGATGATGAAGAAGTG---AAAGAAATGTGCGATGAA---------------------- 405
                 :::   :::||||||            :::|||||||||
Db        436 LeuSerGluGluGluLysTrpSerAlaIleCysAspGluIleLysGluAlaHisThrArg 455

Qy        406 ----------------------TTATCCATAAAAAACACGGACAATTTAAACAGACAA 441
                                     :::||||||:::   ::::::   |||:::
Db        456 GlyGlnProValLeuValGlyThrIleSerIleGluLysSerGluLysLeuSer------ 473

Qy        442 GCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTA 501
                                          |||   :::      |||   |||        |||
Db        474 --------------------AlaLeuLeuArgThrArgGlyValLysHisGluValLeu 486

Qy        502 ------------------------GCTGTCATTGTTGCGAATGCG--------------- 522
                                       |||:::|||:::|||   |||
Db        487 AsnAlaLysAsnHisAlaArgGluAlaLeuIleIleAlaGluAlaGlyAlaLysGlySer 506

Qy        523 ------GTCGCAAAAACCATTCTAGGGCGTGGT------TTATCGCTTGCGGGCAAT 567
                     :::|||      :::   |||||||||      :::   |||   ||||||
Db        507 ValThrIleAlaThrAsnMetAlaGlyArgGlyThrAspIleLysLeuGlyGlyAsn 525

RESULT 10
BIOF_AQUAE
ID   BIOF_AQUAE     STANDARD;      PRT;   373 AA.
AC   O66875;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   8-amino-7-oxononanoate synthase (EC 2.3.1.47) (AONS) (8-amino-7-
DE   ketopelargonate synthase) (7-keto-8-amino-pelargonic acid synthetase)
DE   (7-KAP synthetase) (L-alanine--pimelyl CoA ligase).
GN   BIOF OR AQ_626.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- CATALYTIC ACTIVITY: 6-carboxyhexanoyl-CoA + L-alanine = 8-amino-7-
CC       oxononanoate + CoA + CO(2).
CC   -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC   -!- PATHWAY: Biotin biosynthesis; first step.
CC   -!- SIMILARITY: Belongs to class-II of pyridoxal-phosphate-dependent
CC       aminotransferases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000699; AAC06836.1; -.
DR   PIR; G70355; G70355.
DR   HSSP; P12998; 1BS0.
DR   InterPro; IPR003408; Ala_synthase.
DR   InterPro; IPR004839; Aminotrans_I/II.
DR   InterPro; IPR001917; Aminotrans_II.
DR   InterPro; IPR004723; BioF.
DR   Pfam; PF02490; ALA_synthase; 1.
DR   Pfam; PF00155; aminotran_1_2; 1.
DR   TIGRFAMs; TIGR00858; bioF; 1.
DR   PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
KW   Biotin biosynthesis; Transferase; Pyridoxal phosphate;
KW   Complete proteome.
FT   BINDING     223    223       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   373 AA;  42532 MW;  536B34A5D5F84401 CRC64;

Alignment Scores:
Pred. No.:              2.55          Length:          373
Score:                  96.50         Matches:         64
```

```
Percent Similarity:     36.75%          Conservative:  40
Best Local Similarity:  22.61%          Mismatches:    102
Query Match:            7.15%           Indels:        77
DB:                     1               Gaps:          14

US-09-732-091-3 (1-759) x BIOF_AQUAE (1-373)

Qy         25 TTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTT 84
Db         23 LeuGluGlyValLysAspPheCysSerAsnAspTyrLeu---------------------- 35

Qy         85 TTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATACAAA 144
Db         36 ------------GlyLeuArgLysHisProGluValValGluGluSerIleArgValLeu 51

Qy        145 AGGCATGGC-----------------------------GATGATTACGCTAAATACGCA 174
Db         52 LysGluAlaGlyLeuGlySerGlyAlaSerGlnLeuValSerGlyTyrThrLysHisHis 71

Qy        175 GAAAGAATCGCTGAAGAGTTG---CAATACTATGGGAGCAATAGTTTTGCGAGTTTCATT 231
Db         72 ArgGluLeuGluGluLysLeuAlaGluPheLysGlyThrGluSerCysValLeuPheGly 91

Qy        232 AAAGGC---------------------------------------GAAGGAGTCTTATAC 252
Db         92 SerGlyPheLeuAlaAsnValGlyThrIleProAlaLeuValGluGluGlyAspLeu--- 110

Qy        253 AAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAA 312
Db        111 ------ValLeuSerAsp--------------GluLeuAsnHis--------------- 118

Qy        313 ACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATG 372
Db        119 AlaSerIleIleAspGlyValArgLeuSerLysAlaGlnLysArgValPheLysHisLys 138

Qy        373 GATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAACACGGACAATTTA 432
Db        139 AspTyrGluGluLeuGluGluPheLeuLys-----------LysAsnArgLysLysPhe 154

Qy        433 AACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGT--------- 483
Db        155 ArgArgValLeuIle------IleThrAspThrValPheSerMetAspGlyAspValAla 172

Qy        484 ------------------------TTTAAATCTTATCAATTAGCTGTCATTGTTGCGAAT 519
Db        173 AspLeuLysArgLeuThrGlnIleCysGluGluTyrAspCysMetLeuTyrIleAspGlu 192

Qy        520 GCGGTCGCAAAAACCATTCTAGGGCGT---GGTTTATCGCTTGCGGGCAATCAG------ 570
Db        193 AlaHisThrThrGlyThrIleGlyLysGlyGlyLeuAspTyrPheGlyIleGluHisLys 212

Qy        571 ---------GTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATT 621
Db        213 GluTyrIleIleValMetGlyThrLeuSerLysAlaLeuGlySerTyrGlyAlaPheVal 232

Qy        622 ACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCA 681
Db        233 CysGlyThrLysLeuLeuIleAsp--------------TyrLeuValAsnLysAlaArg 247

Qy        682 TGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAAGCCAATGGAGATAAGAAGTCG 741
Db        248 SerLeuIlePheSerThr---SerLeuProProSerValCysAlaGlyAlaLysLysAla 266

Qy        742 TTGCAAATA 750
Db        267 IleGluIle 269

RESULT 11
ABIC_LACLA
ID   ABIC_LACLA     STANDARD;      PRT;   344 AA.
AC   Q01457;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-1993 (Rel. 27, Last annotation update)
DE   Abortive phage resistance protein abiC.
GN   ABIC OR PRF.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG   Plasmid pTN20.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ME2;
RX   MEDLINE=93054365; PubMed=1429469;
RA   Durmaz E., Higgins D.L., Klaenhammer T.R.;
RT   "Molecular characterization of a second abortive phage resistance
RT   gene present in Lactococcus lactis subsp. lactis ME2.";
RL   J. Bacteriol. 174:7463-7469(1992).
CC   -!- FUNCTION: PROVIDES RESISTANCE TO BACTERIOPHAGE BY ABORTIVE
CC       INFECTION.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M95956; AAA53569.1; -.
DR   PIR; A47025; A47025.
KW   Plasmid.
SQ   SEQUENCE   344 AA;  40128 MW;  AA6E41E71B4A003C CRC64;

Alignment Scores:
Pred. No.:              3.07            Length:        344
Score:                  95.50           Matches:       50
Percent Similarity:     40.95%          Conservative:  36
Best Local Similarity:  23.81%          Mismatches:    73
Query Match:            7.08%           Indels:        51
DB:                     1               Gaps:          9

US-09-732-091-3 (1-759) x ABIC_LACLA (1-344)

Qy         19 AGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGAT---------------------- 57
Db         91 LysAsnGlnGluGluSerLysLysLeuAsnAsnSerAspSerAlaAsnArgGluPheTyr 110

Qy         58 ---TTATTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAAT 114
Db        111 SerLeuLeuAspLeuPheLys--------------LysGluGlnAsnLysSerGluThr 125

Qy        115 GAAAAACTGACCAGCTCCATAGAATACAAAAGG---------------CATGGCGATGAT 159
Db        126 IleLysAlaIleSerPheLeu---TyrLysArgAlaIleAsnAspLysHisGlyAsnSer 144

Qy        160 TACGCTAAA---------------------------------TACGCAGAAAGAATCGCT 186
Db        145 PheIleAsnAspTyrAsnIleAspIleGlyAsnGlyLeuAspPheArgPheAsnLeuPhe 164

Qy        187 GAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTC 246
Db        165 GluSerAsnGlnTrpPheSerIleThrTyrAsnSerSerLysTyrLysGlyLysAsn--- 183

Qy        247 TTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAA 306
Db        184 LeuThrLysGlu-----------------GlnLysValGluLeuValIleSerArgGln 197

Qy        307 ACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAA 366
Db        198 PheAspAspValTyrAsnLysMetSerSerTyrPheLysIlePheHisArgIleLeuLys 217

Qy        367 GAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAACACGGAC 426
Db        218 SerLeuAsnLysArgPheAspGluLysLysLeuAspGluSerAspTyrLysAsnTyrIle 237
```

```
Qy        427 AATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTT 486
                  :::  |||      ||||||
Db        238 GlyIleLeuArgThrGlnLeuSer---------------------------------- 245

Qy        487 AAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTC---GCAAAAACCATTCTAGGG 543
                  |||  :::|||  ||||||:::  |||::::::      |||      ||||||
Db        246 ---SerGluGluLeuValValIleLeuIleAsnSerLeuTyrValLysArgGlyLeuGly 264

Qy        544 CGTGGTTTATCGCTTGCGGGCAATCAGGTG 573
                  |||:::  |||  |||      :::
Db        265 LeuGlyIleGluLeuIleGlyThrAsnLeu 274
```

```
RESULT 12
YC09_CAMJE
ID   YC09_CAMJE     STANDARD;      PRT;   517 AA.
AC   Q9PN86;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical UPF0144 protein Cj1209.
GN   CJ1209.
OS   Campylobacter jejuni.
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Campylobacteraceae; Campylobacter.
OX   NCBI_TaxID=197;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11168;
RX   MEDLINE=20150912; PubMed=10688204;
RA   Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA   Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA   Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA   Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA   Whitehead S., Barrell B.G.;
RT   "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT   reveals hypervariable sequences.";
RL   Nature 403:665-668(2000).
CC   -!- SIMILARITY: Belongs to the UPF0144 family.
CC   -!- SIMILARITY: Contains 1 HD domain.
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL139077; CAB73463.1; -.
DR   PIR; C81327; C81327.
DR   HAMAP; MF_00335; -; 1.
DR   InterPro; IPR006674; HD.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR003607; Met_phsphohydro.
DR   InterPro; IPR006675; Unchar_HDIG.
DR   Pfam; PF01966; HD; 1.
DR   Pfam; PF00013; KH; 1.
DR   SMART; SM00471; HDc; 1.
DR   SMART; SM00322; KH; 1.
DR   TIGRFAMs; TIGR00277; HDIG; 1.
KW   Hypothetical protein; Transmembrane; RNA-binding; Complete proteome.
FT   TRANSMEM      1     21       POTENTIAL.
FT   DOMAIN      206    274       KH.
FT   DOMAIN      333    426       HD.
SQ   SEQUENCE   517 AA;  58025 MW;  6A799602ED363A25 CRC64;
```

```
Alignment Scores:
Pred. No.:              3.41          Length:        517
Score:                 95.00          Matches:        54
Percent Similarity:    38.04%         Conservative:   43
Best Local Similarity: 21.18%         Mismatches:     98
Query Match:            7.04%         Indels:         60
```

```
DB:                    1              Gaps:          10

US-09-732-091-3 (1-759) x YC09_CAMJE (1-517)

Qy          4 GCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTG 63
              |||  :::|||:::  :::|||      ||||||      :::::|||   |||
Db         43 AlaIleGluTyrGluAlaGluLeu---IleLeuLysAspAlaLysAsnSerIleLeuAsn 61

Qy         64 GATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTG 123
                ||||||                  :::|||::::::   ||||||
Db         62 AlaGluLeuGluVal--------------------LysLysLysTyrGluGluLysThr 74

Qy        124 ACCAGC---TCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
                  ::::::  :::      ||||||  :::|||      ::::::
Db         75 HisLysPheGlnLysAspPheAsnGlnLysPheAspAspLeuSerLysLysGluGlnLys 94

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              :::  :::|||  :::          :::
Db         95 LeuGlnGlnGluGluGluLysLeuLysGluAsp---------------------------- 105

Qy        241 GGAGTCTTATACAAAGAGATTTTATGC-------------------------------- 267
                    ||||||  ||||||
Db        106 ------------LysGluTyrLeuCysLysSerGlnLysHisIleGlnAsnLeuGlnSer 121

Qy        268 GATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTA------ 321
              ||||||  ||||||||||||      |||  :::|||  :::      |||
Db        122 AspVal---AspLysLeuLysAsnLysTyrGlnGluLysLeuAspAspValLeuLysIle 140

Qy        322 ------------------ATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTG 363
                              :::  ||||||      :::|||:::  :::  :::
Db        141 LeuGluHisSerThrGlyLeuThrGlnAsnGluAlaLysGluIleIleLeuLysLysVal 160

Qy        364 GAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACG 423
              ||||||          |||::::::      :::      :::          :::
Db        161 GluGluAsnSerArgGluGlnIleAlaHisIleValArgLysTyrGluGluGluAlaLys 180

Qy        424 GACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGT 483
              :::          |||:::|||                                :::      |||
Db        181 AsnGluAlaLysArgLysAlaAsnPheIleIleAlaGlnAlaThrSerArgPheAlaGly 200

Qy        484 TTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGG 543
              :::  :::      :::  |||  |||                :::  |||
Db        201 GluPheAlaAlaGluArgLeuIleAsnValIleAsnIleLysAsnAspGluLeuLysGly 220

Qy        544 CGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGT 603
              |||  :::      |||      |||      :::||||||
Db        221 ArgIleIleGlyLysGluGlyArgAsnVal------LysThrLeuGlu------------ 234

Qy        604 CCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTAT 663
                    ::::::  ||||||      ||||||
Db        235 ------------MetValLeuGlyValAspIleIleIleAsp----------------- 244

Qy        664 AGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTA 708
                    |||      |||:::|||:::      |||
Db        245 ------AspThrProGlyAlaIleIleValSerCysPheAsnLeu 257
```

```
RESULT 13
BPA1_HUMAN
ID   BPA1_HUMAN     STANDARD;      PRT;  3214 AA.
AC   Q03001; Q12825; Q13266; Q13267; Q13775; Q96J76; Q96QT5; Q9UGD7;
AC   Q9UGD8;
DT   01-OCT-1996 (Rel. 34, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Bullous pemphigoid antigen 1 isoforms 1/2/3/4/5/8 (230 kDa bullous
DE   pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein) (Dystonia
DE   musculorum protein) (Fragment).
GN   BPAG1 OR DMH OR DT OR KIAA0728.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
```

```
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   TISSUE=Keratinocytes;
RX   MEDLINE=92011493; PubMed=1717442;
RA   Sawamura D., Li K., Chu M.-L., Uitto J.;
RT   "Human bullous pemphigoid antigen (BPAG1). Amino acid sequences
RT   deduced from cloned cDNAs predict biologically important peptide
RT   segments and protein domains.";
RL   J. Biol. Chem. 266:17784-17790(1991).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   TISSUE=Keratinocytes;
RX   MEDLINE=93346806; PubMed=8345227;
RA   Elgart G.W., Stanley J.R.;
RT   "Cloning of the 5' mRNA for the 230-kD bullous pemphigoid antigen by
RT   rapid amplification of cDNA ends.";
RL   J. Invest. Dermatol. 101:244-246(1993).
RN   [3]
RP   SEQUENCE OF 1-645 FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Fetal brain, and Retina;
RX   MEDLINE=96121394; PubMed=8575775;
RA   Brown A., Dalpe G., Mathieu M., Kothary R.;
RT   "Cloning and characterization of the neural isoforms of human
RT   dystonin.";
RL   Genomics 29:777-780(1995).
RN   [4]
RP   SEQUENCE OF 321-3214 FROM N.A. (ISOFORMS 1 AND 4).
RA   Laird G.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 392-492 FROM N.A.
RC   TISSUE=Pineal gland;
RA   Geerts D., Sonnenberg A.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE OF 1082-3214 FROM N.A. (ISOFORM 3).
RC   TISSUE=Keratinocytes;
RX   MEDLINE=91286285; PubMed=1712022;
RA   Tanaka T., Parry D.A.D., Klaus-Kovtun V., Steinert P.M.,
RA   Stanley J.R.;
RT   "Comparison of molecularly cloned bullous pemphigoid antigen to
RT   desmoplakin I confirms that they define a new family of cell adhesion
RT   junction plaque proteins.";
RL   J. Biol. Chem. 266:12555-12559(1991).
RN   [7]
RP   SEQUENCE OF 2160-2767 FROM N.A.
RX   MEDLINE=91216368; PubMed=2090522;
RA   Owaribe K., Kartenbeck J., Stumpp S., Magin T.M., Krieg T.,
RA   Diaz L.A., Franke W.W.;
RT   "The hemidesmosomal plaque. I. Characterization of a major
RT   constituent protein as a differentiation marker for certain forms of
RT   epithelia.";
RL   Differentiation 45:207-220(1990).
RN   [8]
RP   SEQUENCE OF 2287-3214 FROM N.A. (ISOFORM 5).
RC   TISSUE=Keratinocytes;
RX   MEDLINE=89067122; PubMed=2461961;
RA   Stanley J.R., Tanaka T., Mueller S., Klaus-Kovtun V., Roop D.;
RT   "Isolation of complementary DNA for bullous pemphigoid antigen by use
RT   of patients' autoantibodies.";
RL   J. Clin. Invest. 82:1864-1870(1988).
RN   [9]
RP   SEQUENCE OF 2462-3214 FROM N.A. (ISOFORM 8).
RX   MEDLINE=94280413; PubMed=8010969;
RA   Hopkinson S.B., Jones J.C.;
RT   "Identification of a second protein product of the gene encoding a
RT   human epidermal autoantigen.";
RL   Biochem. J. 300:851-857(1994).
RN   [10]
RP   DOMAINS.
RX   MEDLINE=96199235; PubMed=8621649;
RA   Tang H.-Y., Chaffotte A.-F., Thacher S.M.;
RT   "Structural analysis of the predicted coiled-coil rod domain of the
RT   cytoplasmic bullous pemphigoid antigen (BPAG1). Empirical
RT   localization of the N-terminal globular domain-rod boundary.";
RL   J. Biol. Chem. 271:9716-9722(1996).
RN   [11]
RP   SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND ALTERNATIVE SPLICING.
RC   TISSUE=Keratinocytes;
RX   MEDLINE=21839111; PubMed=11751855;
RA   Okumura M., Yamakawa H., Ohara O., Owaribe K.;
RT   "Novel alternative splicings of BPAG1 (bullous pemphigoid antigen 1)
RT   including the domain structure closely related to MACF (microtubule
RT   actin cross-linking factor).";
RL   J. Biol. Chem. 277:6682-6687(2002).
CC   -!- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
CC       intermediate filaments to the inner plaque of hemidesmosomes. The
CC       proteins may self-aggregate to form filaments or a two-dimensional
CC       mesh.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=10;
CC         Comment=Isoforms 1, 2, 5 and 8 are or may be fragments;
CC       Name=1;
CC         IsoId=Q03001-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q03001-2; Sequence=VSP_005053, VSP_005055;
CC       Name=3; Synonyms=1e;
CC         IsoId=Q03001-3; Sequence=VSP_005054, VSP_005056, VSP_005057,
CC                                  VSP_005058, VSP_005059, VSP_005060,
CC                                  VSP_005061;
CC       Name=4;
CC         IsoId=Q03001-4; Sequence=VSP_005054, VSP_005056;
CC       Name=5;
CC         IsoId=Q03001-5; Sequence=VSP_005064, VSP_005065;
CC       Name=6; Synonyms=EA;
CC         IsoId=O94833-2; Sequence=External;
CC       Name=7; Synonyms=EB;
CC         IsoId=Q8WXK8-1; Sequence=External;
CC       Name=8;
CC         IsoId=Q03001-6; Sequence=VSP_005062, VSP_005063;
CC       Name=9;
CC         IsoId=O94833-3; Sequence=External;
CC       Name=10;
CC         IsoId=O94833-1; Sequence=External;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Highly expressed in skeletal muscle and
CC       cultured keratinocytes.
CC   -!- DISEASE: BPAG1 is an autoantigen of bullous pemphigoid
CC       [MIM:600088], an autoimmune subepithelial skin blistering disease.
CC   -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC   -!- SIMILARITY: Contains 1 actin-binding domain.
CC   -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   -!- SIMILARITY: Contains 11 plectin repeats.
CC   -!- SIMILARITY: Contains 4 spectrin repeats.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M69225; -; NOT_ANNOTATED_CDS.
DR   EMBL; L11690; AAA52288.1; -.
DR   EMBL; U31850; AAC50243.1; -.
DR   EMBL; U31851; AAC50244.1; -.
DR   EMBL; AL096710; -; NOT_ANNOTATED_CDS.
DR   EMBL; AY032900; AAK63130.1; -.
DR   EMBL; AY032901; AAK63131.1; -.
DR   EMBL; M63618; AAA35606.1; -.
DR   EMBL; X58677; CAA41528.1; -.
DR   EMBL; M22942; AAA35538.1; -.
DR   EMBL; U04850; AAA57184.1; -.
```

```
DR   EMBL; U04850; AAA57185.1; -.
DR   PIR; I56317; A40937.
DR   Genew; HGNC:1090; BPAG1.
DR   MIM; 113810; -.
DR   MIM; 600088; -.
DR   GO; GO:0005604; C:basement membrane; TAS.
DR   GO; GO:0005737; C:cytoplasm; IEP.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; IEP.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; TAS.
DR   GO; GO:0045104; P:intermediate filament cytoskeleton organiza. . .; IEP.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR001101; Plectin_repeat.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00681; Plectin; 5.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00435; spectrin; 3.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00150; SPEC; 2.
DR   PROSITE; PS50002; SH3; FALSE_NEG.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR   PROSITE; PS50021; CH; 2.
KW   Antigen; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW   Structural protein; Cytoskeleton; Cell adhesion; Calcium;
KW   Calcium-binding; Alternative splicing.
FT   NON_TER       1     1
FT   DOMAIN        1   1669       GLOBULAR 1.
FT   DOMAIN     1670   2441       CENTRAL FIBROUS ROD DOMAIN.
FT   DOMAIN     2442   3214       GLOBULAR 2.
FT   DOMAIN      215    439       ACTIN-BINDING.
FT   DOMAIN      219    322       CH 1.
FT   DOMAIN      335    436       CH 2.
FT   REPEAT      774    851       SPECTRIN 1.
FT   REPEAT      876    959       SPECTRIN 2.
FT   REPEAT      984   1077       SPECTRIN 3.
FT   DOMAIN     1091   1143       SH3.

Alignment Scores:
Pred. No.:              3.61          Length:        3214
Score:                 95.00          Matches:         46
Percent Similarity:    44.75%         Conservative:    35
Best Local Similarity: 25.41%         Mismatches:      62
Query Match:            7.04%         Indels:          38
DB:                     1             Gaps:             9

US-09-732-091-3 (1-759) x BPA1_HUMAN (1-3214)

Qy          4 GCATACAAATATGATAGAGACTTGGAATTTTTAAAG------CAATTGGAATCTAGTGAT 57
              :::::::::   :::::::::||||||   |||:::      |||   :::|||
Db       2075 SerPheArgAspGluLysGluLeuGluArgLeuGlnIleCysGlnArgLysSerAspHis 2094

Qy         58 TTATTGGATTTGTTT----------GAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAA 105
              |||   :::   |||           |||   ||:::              ||||||
Db       2095 LeuLysGluGlnPheGluLysSerHisGluGlnLeuLeuGlnAsnIleLysAlaGluLys 2114

Qy        106 AGACACAATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCT 165
              :::|||:::||:::           |||              ::::::   |||
Db       2115 GluAsnAsnAspLysIleGlnArgLeuAsnGluGluLeuGluLysSerAsnGluCysAla 2134

Qy        166 AAATACGCAGAAAGAATCGCTGAAGAGTTG------CAATACTATGGGAGCAATAGTTTT 219
              :::      ::::::      |||||||||      |||         :::
Db       2135 GluMetLeuLysGlnLysValGluGluLeuThrArgGlnAsnAsnGluThrLysLeuMet 2154

Qy        220 GCGAGTTTCATTAAAGGCGAAGGA---------GTCTTATACAAAGAGATTTTATGCGAT 270
                 |||:::  |||           ||||||  |||:::  :::
Db       2155 MetGlnArgIleGlnAlaGluSerGluAsnIleValLeuGluLysGlnThrIleGlnGln 2174

Qy        271 GTGTGCGATAAATTAAAGGTC------------------------------------ 291
                |||:::  ||||||:::
Db       2175 ArgCysGluAlaLeuLysIleGlnAlaAspGlyPheLysAspGlnLeuArgSerThrAsn 2194

Qy        292 ---AATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATC 348
                 :::   :::|||:::|||:::|||          ||||||:::         ||||||
Db       2195 GluHisLeuHisLysGlnThrLysThr---------GluGlnAspPheGlnArgLysIle 2211

Qy        349 ------TTAGAAAGAAGTTTGGAAGAAATGGATGAT------GAAGAAGTGAAAGAAATG 396
                    ||||||      |||   :::      :::          |||   |||:::
Db       2212 LysCysLeuGluGluAspLeuAlaLysSerGlnAsnLeuValSerGluPheLysGlnLys 2231

Qy        397 TGCGATGAATTATCCATA-----------------AAAAACACGGACAATTTAAACAGA 438
              ||||||:::   :::|||                 |||         |||||||||
Db       2232 CysAspGlnGlnAsnIleIleIleGlnAsnThrLysLysGluValArgAsnLeuAsnAla 2251

Qy        439 CAA 441
              :::
Db       2252 Glu 2252

RESULT 14
REST_CHICK
ID   REST_CHICK     STANDARD;      PRT;  1433 AA.
AC   O42184; O42228; O57563; O57564;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Restin (Cytoplasmic linker protein-170) (CLIP-170).
GN   RSN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98137792; PubMed=9469933;
RA   Griparic L., Volosky J.M., Keller T.C. III;
RT   "Cloning and expression of chicken CLIP-170 and restin isoforms.";
RL   Gene 206:195-208(1998).
RN   [2]
RP   SEQUENCE OF 17-1139 FROM N.A. (ISOFORMS 3 AND 4).
RC   TISSUE=Pectoralis muscle;
RX   MEDLINE=99002898; PubMed=9784600;
RA   Griparic L., Keller T.C. III;
RT   "Identification and expression of two novel CLIP-170/Restin isoforms
RT   expressed predominantly in muscle.";
RL   Biochim. Biophys. Acta 1405:35-46(1998).
CC   -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC       THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC       CYTOSKELETON (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=O42184-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O42184-2; Sequence=VSP_000761;
CC       Name=3; Synonyms=CLIP-170(11);
CC         IsoId=O42184-3; Sequence=VSP_000762, VSP_000763;
CC       Name=4; Synonyms=CLIP-170(11+35);
CC         IsoId=O42184-4; Sequence=VSP_000764;
CC   -!- SIMILARITY: Contains 2 CAP-Gly domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF014012; AAC60344.1; -.
DR   EMBL; AF020764; AAC60345.1; -.
DR   EMBL; AF045650; AAC03547.1; -.
```

```
DR   EMBL; AF045651; AAC03548.1; -.
DR   InterPro; IPR000938; CAP-Gly.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01302; CAP_GLY; 2.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS00845; CAP_GLY_1; 2.
DR   PROSITE; PS50245; CAP_GLY_2; 2.
KW   Cytoskeleton; Microtubule; Coiled coil; Repeat; Alternative splicing.
FT   DOMAIN       79    121       CAP-GLY 1.
FT   DOMAIN      144    207       SER-RICH.
FT   DOMAIN      235    277       CAP-GLY 2.
FT   DOMAIN      305    332       SER-RICH.
FT   DOMAIN      351   1353       COILED COIL (POTENTIAL).
FT   DOMAIN     1414   1427       CCHC-BOX.
FT   VARSPLIC    458    492       Missing (in isoform 2).
FT                                /FTId=VSP_000761.
FT   VARSPLIC    458    492       TQTKLEHARIKELEQSLLFEKTKADKLQRELEDTR -> RK
FT                                RQISEDPEN (in isoform 3).
FT                                /FTId=VSP_000762.
FT   VARSPLIC    803    803       S -> GGSSKVS (in isoform 3).
FT                                /FTId=VSP_000763.
FT   VARSPLIC    458    458       T -> RKRQISEDPENT (in isoform 4).
FT                                /FTId=VSP_000764.
FT   CONFLICT    309    309       K -> R (IN REF. 2; AAC03547).
FT   CONFLICT    440    440       E -> V (IN REF. 2; AAC03548).
SQ   SEQUENCE   1433 AA;  161026 MW;  5631CE8683498E23 CRC64;

Alignment Scores:
Pred. No.:                3.86          Length:        1433
Score:                    94.50         Matches:       52
Percent Similarity:       41.31%        Conservative:  36
Best Local Similarity:    24.41%        Mismatches:    74
Query Match:              7.01%         Indels:        51
DB:                       1             Gaps:          10

US-09-732-091-3 (1-759) x REST_CHICK (1-1433)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              |||      |||:::|||   |||::::::      ||||||           :::|||
Db       1020 MetIleThrLysHisAspAlaAspIleLysGlyPheLysGln-----------AsnLeu 1035

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||      |||   |||         |||      :::|||::::::::::|||
Db       1036 LeuAspAlaGluGluAlaLeu---------LysAlaAlaGlnLysLysAsnAspGluLeu 1052

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
               |||   :::   |||   |||:::   :::      ||||||      :::|||
Db       1053 GluThrGlnAlaGluGluLeuLysLysGlnAlaGluGln---AlaLysAlaAspLysArg 1071

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
                 |||   ||||||              :::                 |||   :::
Db       1072 AlaGluGluValLeuGln--------------ThrMetGluLysValThrLysGluLys 1086

Qy        241 GGAGTCTTATACAAA---------GAGATTTTATGCGATGTGTGCGATAAATTAAAGGTC 291
                    :::::::::         |||   |||      :::   :::      :::
Db       1087 AspAlaIleHisGlnGluLysIleGluThrLeuAlaSerLeuGluAsnSerArgGlnThr 1106

Qy        292 AATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTA 351
              |||      :::   |||      :::::::::|||||   |||
Db       1107 AsnGluLysLeuGlnAsnGluLeuAspMetLeuLysGlnAsnAsnLeuLysAsnGluGlu 1126

Qy        352 GAA-----------AGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGC 399
              |||          :::   |||      :::::::::::::::|||:::|||:::
Db       1127 GluLeuThrLysSerLysGluLeuLeuAsnLeuGluAsnLysLysValGluGluLeuLys 1146

Qy        400 GATGAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACT 459
                 |||                                  :::||||||      |||
Db       1147 LysGluPhe-----------------------------GluAlaLeuLysLeuAlaAla 1156

Qy        460 TTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATT---GTTGCG 516
                                              ||||||   |||||||||   :::
Db       1157 AlaGln--------------------LysSerGlnGlnLeuAlaAlaLeuGlnGluGlu 1169

Qy        517 AATGCGGTCGCAAAAACCATTCTAGGGCGTGGT---------------------------- 549
              |||                        |||||||||
Db       1170 AsnValLysLeuAlaGluGluLeuGlyArgSerArgAspGluValThrSerHisGlnLys 1189

Qy        550 ------------TTATCGCTTGCGGGCAATCAGGTGCTT 576
                          |||:::      ||||||:::|||
Db       1190 LeuGluGluGluArgSerValLeuAsnAsnGlnLeuLeu 1202

RESULT 15
PAC_BACME
ID   PAC_BACME      STANDARD;      PRT;   802 AA.
AC   Q60136; Q9S463;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
DE   (Penicillin G amidohydrolase).
GN   PAC OR PGA.
OS   Bacillus megaterium.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1404;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 14945;
RA   Kang J.H., Kim S.J., Park Y.C., Hwang Y., Yoo O.J., Kim Y.C.;
RT   "Nucleotide sequence of the penicillin G acylase gene from
RT   Bacillus megaterium and characteristics of the enzyme.";
RL   Misainmurhag Hoiji 32:215-221(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 14945;
RX   MEDLINE=95180705; PubMed=7875576;
RA   Martin L.M., Prieto A.M., Cortes E., Garcia J.L.;
RT   "Cloning and sequencing of the pac gene encoding the penicillin G
RT   acylase of Bacillus megaterium ATCC 14945.";
RL   FEMS Microbiol. Lett. 125:287-292(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CA4098;
RX   PubMed=12114980;
RA   Yang S., Huang Y.H., Huang X.D., Li S.Y., Yuan Z.Y.;
RT   "High expression of penicillin G acylase gene from Bacillus megaterium
RT   in Bacillus subtilis.";
RL   Acta Biochim. Biophys. Sin. 31:601-603(1999).
CC   -!- CATALYTIC ACTIVITY: Penicillin + H(2)O = a fatty acid anion + 6-
CC       aminopenicillanate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (Potential).
CC   -!- SUBUNIT: Heterodimer of an alpha chain and a beta chain (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Extracellular (Potential).
CC   -!- SIMILARITY: Belongs to peptidase family S45.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U07682; AAB41343.1; -.
DR   EMBL; Z37542; CAA85774.1; -.
DR   EMBL; AF161313; AAD45609.1; -.
DR   PIR; S49252; S49252.
DR   HSSP; P06875; 1AJQ.
DR   MEROPS; S45.001; -.
DR   InterPro; IPR002692; Peptidase_S45.
DR   Pfam; PF01804; Penicil_amidase; 1.
KW   Hydrolase; Antibiotic resistance; Zymogen; Calcium-binding; Signal.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25    802       PENICILLIN G ACYLASE ZYMOGEN.
FT   CHAIN       25    234       PENICILLIN G ACYLASE ALPHA SUBUNIT.
FT   PROPEP     235    265       SPACER PEPTIDE.
```

```
FT   CHAIN       266    802       PENICILLIN G ACYLASE BETA SUBUNIT.
FT   ACT_SITE    266    266       BY SIMILARITY.
FT   METAL       177    177       CALCIUM (POTENTIAL).
FT   METAL       341    341       CALCIUM (POTENTIAL).
FT   VARIANT       3      3       T -> M (IN STRAIN CA4098).
FT   VARIANT     224    224       D -> E (IN STRAIN CA4098).
FT   VARIANT     232    232       I -> K (IN STRAIN CA4098).
FT   VARIANT     254    254       T -> S (IN STRAIN CA4098).
FT   VARIANT     349    349       A -> T (IN STRAIN CA4098).
FT   VARIANT     470    470       S -> N (IN STRAIN CA4098).
FT   VARIANT     524    524       F -> Y (IN STRAIN CA4098).
FT   VARIANT     569    569       Q -> P (IN STRAIN CA4098).
FT   VARIANT     586    586       I -> A (IN STRAIN CA4098).
FT   VARIANT     657    657       N -> S (IN STRAIN CA4098).
FT   VARIANT     740    740       T -> K (IN STRAIN CA4098).
FT   VARIANT     789    791       NKA -> YKS (IN STRAIN CA4098).
SQ   SEQUENCE   802 AA;  91987 MW;  877CA0564E50DFBD CRC64;
```

Alignment Scores:

Pred. No.:	4.57	Length:	802
Score:	93.50	Matches:	62
Percent Similarity:	35.44%	Conservative:	39
Best Local Similarity:	21.75%	Mismatches:	107
Query Match:	6.93%	Indels:	77
DB:	1	Gaps:	10

US-09-732-091-3 (1-759) x PAC_BACME (1-802)

```
Qy         10 AAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTG 69
Db         33 LysValValArgAspAsnPheGlyValProHisLeuTyrAlaLysAsnLysLysAspLeu 52

Qy         70 TTTGAGGTG------CTTGTTTTTGGTAAAGACGGC----------------GAAAAA 105
Db         53 TyrGluAlaTyrGlyTyrValMetAlaLysAspArgLeuPheGlnLeuGluMetPheArg 72

Qy        106 AGACACAATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCT 165
Db         73 ArgGlyAsnGluGlyThrValSerGluIle------------PheGlyGluAspTyrLeu 88

Qy        166 ------------------------------------AAATACGCAGAAAGAATCGCT 186
Db         89 SerLysAspGluGlnSerArgArgAspGlyTyrSerAsnLysGluIleLysLysMetIle 108

Qy        187 GAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTC 246
Db        109 AspGlyLeuAspArgGlnProLysGluLeuIleAlaLysPheAlaGluGlyIleSerArg 128

Qy        247 TTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC------ 300
Db        129 TyrValAsnGluAlaLeuLysAspProAspAspLysLeuSerLysGluPheHisGluTyr 148

Qy        301 --------------AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCT--- 342
Db        149 GlnPheLeuProGlnLysTrpThrSerThrAspValValArgValTyrMetValSerMet 168

Qy        343 ------------------------------------AAAATCTTAGAAAGAAGTTTG 363
Db        169 ThrTyrPheMetAspAsnHisGlnGluLeuLysAsnAlaGluIleLeuAlaLysLeuGlu 188

Qy        364 GAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC--- 420
Db        189 HisGluTyrGlyThrGluValSerArgLysMetPheAspAspLeuValTrpLysAsnAsp 208

Qy        421 ------------------------------------------------ACGGACAATTTA 432
Db        209 ProSerAlaProThrSerIleValSerGluGlyLysProLysArgAspSerSerSerGln 228

Qy        433 AACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAA--- 489
Db        229 SerLeuGlnIleLeuSerSerAlaValIleLysAlaSerGluLysValGlyLysGluArg 248

Qy        490 ------------------TCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAA 531
```

```
Db        249 GluAsnPheValGlnThrSerGluGluLeuGlyLeuProLeuLysIleGlySerAsnAla 268

Qy        532 ACCATTCTAGGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGC 591
Db        269 AlaIleValGlySerGluLysSerAlaThrGlyAsnAlaLeuLeu---------------- 283

Qy        592 TTTTTAACAGGTCCT---GTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATT 648
Db        284 ---PheSerGlyProGlnValGlyPheValAlaProGlyPheLeuTyrGluValGlyLeu 302

Qy        649 GCAGGGCCGGCTTAT 663
Db        303 HisAlaProGlyPhe 307
```

Search completed: July 5, 2004, 03:06:43
Job time : 20.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 5, 2004, 02:09:47 ; Search time 80.5 Seconds
(without alignments)
5949.772 Million cell updates/sec

Title: US-09-732-091-3
Perfect score: 1349
Sequence: 1 atggcatacaaatatgatag.....cgttgcaaatagaatccatt 759

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09732091/runat_02072004_182407_19014/app_query.fasta_1.903
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09732091_@CGN_1_1_71_@runat_02072004_182407_19014 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	265	19.6	237	16	Q8XGV3	Q8xgv3 salmonella
	2	260	19.3	237	16	Q83SR4	Q83sr4 shigella fl
	3	259	19.2	237	16	Q8FLC7	Q8flc7 escherichia
	4	133	9.9	39	16	O26108	O26108 helicobacte
c	5	127.5	9.5	206	16	Q8XA70	Q8xa70 escherichia
c	6	125	9.3	199	16	Q7UDU2	Q7udu2 shigella fl
	7	105	7.8	1245	5	Q8IAN1	Q8ian1 plasmodium
	8	103.5	7.7	403	16	Q8R9Z4	Q8r9z4 thermoanaer
	9	102.5	7.6	276	16	Q8DIH4	Q8dih4 synechococc
	10	102	7.6	1590	5	Q8MX30	Q8mx30 naegleria g
	11	101	7.5	1455	5	Q8IKG8	Q8ikg8 plasmodium
	12	100.5	7.4	726	13	P87397	P87397 *oncorhynchu*
	13	99.5	7.4	556	16	Q7VHM0	Q7vhm0 helicobacte
	14	99.5	7.4	633	16	Q882L0	Q88210 pseudomonas
	15	99	7.3	202	5	Q8IJZ2	Q8ijz2 plasmodium
	16	99	7.3	222	4	O43633	O43633 homo sapien
	17	99	7.3	222	11	Q9DB34	Q9db34 mus musculu
	18	99	7.3	723	13	Q7ZZU9	Q7zzu9 astyanax fa
	19	98.5	7.3	1009	16	Q99UD0	Q99ud0 staphylococ
	20	98	7.3	282	16	Q8ZCL9	Q8zcl9 yersinia pe
	21	98	7.3	413	9	Q8SDW1	Q8sdw1 bacteriopha
	22	98	7.3	1099	16	Q98QT2	Q98qt2 mycoplasma
	23	98	7.3	2359	5	Q8I297	Q8i297 plasmodium
	24	98	7.3	2849	5	Q8IHY4	Q8ihy4 plasmodium
	25	97.5	7.2	311	16	Q9PIC9	Q9pic9 campylobact
	26	97.5	7.2	728	16	Q9PI94	Q9pi94 campylobact
	27	97.5	7.2	792	10	Q9ARQ9	Q9arq9 oryza sativ
c	28	97	7.2	346	16	O51329	O51329 borrelia bu
	29	96.5	7.2	1102	16	Q827Q7	Q827q7 streptomyce
	30	96.5	7.2	1341	5	Q8IEQ5	Q8ieq5 plasmodium
	31	96	7.1	262	5	O61768	O61768 caenorhabdi
	32	96	7.1	420	16	Q97MM5	Q97mm5 clostridium
	33	96	7.1	4405	5	Q8ILZ2	Q8ilz2 plasmodium
	34	95.5	7.1	944	17	Q8TPC2	Q8tpc2 methanosarc
	35	94.5	7.0	387	10	Q9FX81	Q9fx81 arabidopsis
	36	94.5	7.0	406	16	Q8RH57	Q8rh57 fusobacteri
	37	94	7.0	385	2	Q9ZI88	Q9zi88 borrelia bu
	38	94	7.0	576	10	Q84LB2	Q84lb2 malus domes
c	39	93.5	6.9	408	10	Q94JW4	Q94jw4 oryza sativ
	40	93.5	6.9	681	10	Q9XIK2	Q9xik2 *arabidopsis*
c	41	93.5	6.9	701	17	Q973X0	Q973x0 sulfolobus
	42	93.5	6.9	1413	5	Q8I589	Q8i589 plasmodium
	43	93	6.9	355	17	Q97UF7	Q97uf7 *sulfolobus*
	44	93	6.9	610	16	O51577	O51577 borrelia bu
c	45	93	6.9	667	11	O35267	O35267 rattus norv

ALIGNMENTS

RESULT 1
Q8XGV3
ID Q8XGV3 PRELIMINARY; PRT; 237 AA.
AC Q8XGV3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Positive regulator for sigma H (Sigma 32)
DE promoters, permitting growth at high temperature).
GN T0010 OR HTGA OR STM0010 OR STY0010.
OS Salmonella typhi, and
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601, 602;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
RN [2]

```
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.typhi; STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AE016834; AAO67744.1; -.
DR   EMBL; AE008693; AAL18974.1; -.
DR   EMBL; AL627265; CAD01163.1; -.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
XW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   237 AA;  26515 MW;  184ADE026EA5BCA9 CRC64;

Alignment Scores:
Pred. No.:                7.98e-13        Length:          237
Score:                    265.00          Matches:         79
Percent Similarity:       49.17%          Conservative:    40
Best Local Similarity:    32.64%          Mismatches:      101
Query Match:              19.64%          Indels:          22
DB:                       16              Gaps:            7

US-09-732-091-3 (1-759) x Q8XGV3 (1-237)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              :::   |||   :::|||   ||||||:::||||||:::                  |||
Db          3 ValThrTyrLeuHisAspGluAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeu 22

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGA---------- 108
                |||        :::|||        :::   |||:::   |||
Db         23 AlaAspPheAlaArgLeuLeuThrHisAsnGluLysGlyLysAlaArgLeuSerServal 42

Qy        109 ------CACAATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTAC 162
                   |||||||||   |||      ::::::|||         |||   :::   :::
Db         43 LeuSerHisAsnGlu---LeuPheLysAlaMetGlu------GlyHisProGluGlnHis 59

Qy        163 GCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCG 222
              :::      :::   |||||   |||   |||:::||||||   :::|||   |||
Db         60 ArgArgAsnTrpGlnLeuIleAlaGlyGluPheGlnHisTyrGlyGlyAspSerIleAla 79

Qy        223 AGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAA 282
              :::   ::::::|||   |||      |||:::   ||||||   ||||||      :::
Db         80 AsnLysLeuArgGlyHisGlyLysGlnTyrArgAlaIleLeuLeuAspValAlaLysArg 99

Qy        283 TTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCT 342
              ||||||:::      :::|||          |||      |||||||||   :::|||
Db        100 LeuLysLeuLysAlaAspLysSerMetSerThrPheGluIleGluGlnGlnLeuLeuGlu 119

Qy        343 AAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGAT 402
                  |||      :::   ::::::||||||            :::|||
Db        120 HisPheLeuArgHisThrTrpGlnLysMetAspAlaAlaHisLysGlnGluPheLeuGln 139

Qy        403 GAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTA 462
                 :::      |||   ::::::          |||
Db        140 AlaValAspAlaLysValSerGluLeuGluGluLeuProLeuLeuMetLysAspArg 159

Qy        463 ACGCTGTTTAAA---ATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAAT 519
              :::|||   |||   :::           |||   ||||||      |||:::
Db        160 SerLeuAlaLysGlyValSerHisLeuLeuSerThrGlnLeuThrArgIleLeuArgThr 179

Qy        520 GCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACA 579
                  |||   :::|||||||||   ||||||   |||   |||
Db        180 HisAlaAlaMetSerIleLeuGlyHisGlyLeu---LeuArgGly--------------- 193

Qy        580 AGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCG 639
                    :::   |||   |||||||||||       :::   ||||||
Db        194 ---------AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsnGlyValLysAla--- 209

Qy        640 ATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACT 699
                  :::::::|||   |||||||||||||||||||||||   :::   :::|||
Db        210 ------MetSerGlySerAlaTyrArgValThrIleProAlaValLeuGlnIleAlaCys 227

Qy        700 TTACGC 705
                ||||||
Db        228 LeuArg 229

RESULT 2
Q83SR4
ID   Q83SR4      PRELIMINARY;      PRT;   237 AA.
AC   Q83SR4;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative oxidoreductase.
GN   SF0012 OR S0012.
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Shigella.
OX   NCBI_TaxID=623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=301 / Serotype 2a;
RX   MEDLINE=22272406; PubMed=12384590;
RA   Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA   Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA   Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA   Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA   Yu J.;
RT   "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT   through comparison with genomes of Escherichia coli K12 and O157.";
RL   Nucleic Acids Res. 30:4432-4441(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX   MEDLINE=22590274; PubMed=12704152;
RA   Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA   Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA   Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA   Schwartz D.C., Blattner F.R.;
RT   "Complete genome sequence and comparative genomics of Shigella
RT   flexneri serotype 2a strain 2457T.";
RL   Infect. Immun. 71:2775-2786(2003).
DR   EMBL; AE015039; AAN41678.1; -.
DR   EMBL; AE016978; AAP15557.1; -.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Complete proteome.
SQ   SEQUENCE   237 AA;  26709 MW;  95509A1BCB8B4CF5 CRC64;

Alignment Scores:
Pred. No.:                1.99e-12        Length:          237
Score:                    260.00          Matches:         76
Percent Similarity:       48.58%          Conservative:    44
Best Local Similarity:    30.77%          Mismatches:      91
Query Match:              19.27%          Indels:          36
DB:                       16              Gaps:            6
```

```
US-09-732-091-3 (1-759) x Q83SR4 (1-237)

Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
              |||       |||   ||||||:::||||||:::                  |||   :::
Db          5 TyrLeuAsnAspSerAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeuAlaAsn 24

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
                  :::|||          :::   |||:::   |||   :::       |||
Db         25 PheAlaArgLeuLeuThrHisAsnGluLysGlyLysThrArgLeuSerSerIleLeuXet 44

Qy        127 AGCTCCATAGAATACAAAAGG---------CATGGCGATGATTACGCTAAATACGCAGAA 177
                 :::       :::|||            |||   :::   :::   :::       :::
Db         45 ArgAsnGluLeuPheLysSerMetGluGlyHisProGluGlnHisArgArgAsnTrpGln 64

Qy        178 AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGC 237
              ||||||   |||||||:::::|||   :::||   |||:::   ::::::|||
Db         65 LeuIleAlaGlyGluLeuGlnHisPheGlyGlyAspSerIleAlaAsnLysLeuArgGly 84

Qy        238 GAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTAC 297
              |||   |||||:::   ||||||   ||||||       :::|||||:::
Db         85 HisGlyLysLeuTyrArgAlaIleLeuLeuAspValSerLysArgLeuLysLeuLysAla 104

Qy        298 AACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGA 357
              :::|||:::       |||       |||||||||   :::|||   :::   |||
Db        105 AspLysGluMetSerThrPheGluIleGluGlnGlnLeuLeuGluGlnPheLeuArgAsn 124

Qy        358 AGTTTGGAAGAAATGGATGATGAA-------------------------------GAA 384
              :::   ::::::||||||:::|||
Db        125 ThrTrpLysLysMetAspGluGluHisLysGlnGluPheLeuHisAlaValAspAlaArg 144

Qy        385 GTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCC 444
              |||   |||:::   :::   |||   :::         |||   |||   :::
Db        145 ValAsnGluLeuGluGluLeuLeuProLeuLeuMetLysAspLysLeuLeuAlaLysGly 164

Qy        445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCT 504
              :::|||                                        |||   ||||||
Db        165 ValSer-------------------------------HisLeuLeuSerSerGlnLeuThr 174

Qy        505 GTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGC 564
              |||:::           |||   ::::::|||||   ||||||   |||   |||
Db        175 ArgIleLeuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu---LeuArgGly 193

Qy        565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACA 624
                          :::   |||   ||||||||||||       :::
Db        194 ------------------------AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205

Qy        625 GGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGC 684
              ||||||                 ::::::|||      ||||||||||||||||||||
Db        206 GlyValLysAla---------ValSerGlySerThrTyrArgValThrIleProAlaVal 222

Qy        685 ATTGTGGTTGCCACTTTACGC 705
              :::   :::|||   ||||||
Db        223 LeuGlnIleAlaCysLeuArg 229

RESULT 3
Q8FLC7
ID   Q8FLC7      PRELIMINARY;      PRT;   237 AA.
AC   Q8FLC7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein yaaW.
GN   YAAW OR C0016.
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR   EMBL; AE016755; AAN78516.1; -.
DR   InterPro; IPR005367; UPF0174.
DR   Pfam; PF03667; UPF0174; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   237 AA;  26727 MW;  B8C190712375B31D CRC64;

Alignment Scores:
Pred. No.:                2.39e-12      Length:          237
Score:                    259.00        Matches:         76
Percent Similarity:       48.58%        Conservative:    44
Best Local Similarity:    30.77%        Mismatches:      91
Query Match:              19.20%        Indels:          36
DB:                       16            Gaps:            6

US-09-732-091-3 (1-759) x Q8FLC7 (1-237)

Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
              |||       |||   ||||||:::||||||:::                  |||   :::
Db          5 TyrLeuAsnAspSerAspLeuAspPheLeuGlnHisCysSerGluGluGlnLeuAlaAsn 24

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
                  :::|||          :::   |||:::   |||   :::       |||
Db         25 PheAlaArgLeuLeuThrHisAsnGluLysGlyLysThrArgLeuSerSerValLeuMet 44

Qy        127 AGCTCCATAGAATACAAAAGG---------CATGGCGATGATTACGCTAAATACGCAGAA 177
                 :::       :::|||            |||   :::   :::   :::       :::
Db         45 ArgAsnGluLeuPheLysSerMetGluGlyHisProGluArgHisArgArgAsnTrpGln 64

Qy        178 AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGC 237
              ||||||   |||   |||:::::|||   :::|||   |||:::   ::::::|||
Db         65 LeuIleAlaGlyGluPheGlnHisPheGlyGlyAspSerIleAlaAsnLysLeuArgGly 84

Qy        238 GAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTAC 297
              |||   |||||:::   ||||||   ||||||       :::|||||:::
Db         85 HisGlyLysLeuTyrArgAlaIleLeuLeuAspValSerLysArgLeuLysLeuLysAla 104

Qy        298 AACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGA 357
              :::|||:::       |||       |||||||||   :::|||   :::   |||
Db        105 AspLysGluMetSerThrPheGluIleGluGlnGlnLeuLeuGluGlnPheLeuArgAsn 124

Qy        358 AGTTTGGAAGAAATGGATGATGAA-------------------------------GAA 384
              :::   ::::::||||||:::|||
Db        125 ThrTrpLysLysMetAspGluGluHisLysGlnGluPheLeuHisAlaValAspAlaArg 144

Qy        385 GTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCC 444
              |||   |||:::   :::   |||   :::         |||   |||   :::
Db        145 ValAsnGluLeuGluGluLeuLeuProLeuLeuMetLysAspLysLeuLeuAlaLysGly 164

Qy        445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCT 504
              :::|||                                        |||   ||||||
Db        165 ValSer-------------------------------HisLeuLeuSerSerGlnLeuThr 174

Qy        505 GTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGC 564
              |||:::           |||   ::::::|||||   ||||||   |||   |||
Db        175 ArgIleLeuArgThrHisAlaAlaMetSerValLeuGlyHisGlyLeu---LeuArgGly 193

Qy        565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACA 624
                          :::   |||   ||||||||||||       :::
Db        194 ------------------------AlaGlyLeuGlyGlyProValGlyAlaAlaLeuAsn 205

Qy        625 GGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGC 684
              ||||||                 ::::::|||   ||||||||||||||||||||
Db        206 GlyValLysAla---------ValSerGlySerAlaTyrArgValThrIleProAlaVal 222

Qy        685 ATTGTGGTTGCCACTTTACGC 705
              :::   :::|||   ||||||
Db        223 LeuGlnIleAlaCysLeuArg 229
```

```
RESULT 4
O26108
ID   O26108      PRELIMINARY;      PRT;    39 AA.
AC   O26108;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein HP1590.
GN   HP1590.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
DR   EMBL; AE000656; AAD08629.1; -.
DR   PIR; F64718; F64718.
DR   TIGR; HP1590; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   39 AA;  4636 MW;  EE53CF01F4C1C909 CRC64;

Alignment Scores:
Pred. No.:                0.0235          Length:         39
Score:                    133.00          Matches:        26
Percent Similarity:       79.49%          Conservative:   5
Best Local Similarity:    66.67%          Mismatches:     8
Query Match:              9.86%           Indels:         0
DB:                       16              Gaps:           0

US-09-732-091-3 (1-759) x O26108 (1-39)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              |||||||||:::||||||   ||||||||||||||||:::|||   ||||||||||||
Db          1 MetAlaTyrArgTyrAspSerAspLeuGluPheLeuLysArgLeuSerSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAA 117
                |||||||||:::   ||||||:::   :::||||||      |||   ||||||
Db         21 LysAspLeuPheAspAlaLeuValTyrAspGluAspGlyThrLeuArgMetAsnGlu 39

RESULT 5
Q8XA70
ID   Q8XA70      PRELIMINARY;      PRT;   206 AA.
AC   Q8XA70;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Positive regulator for sigma 32 heat shock promoters.
GN   HTGA OR Z0012 OR ECS0011.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005178; AAG54312.1; -.
DR   EMBL; AP002550; BAB33435.1; -.
DR   PIR; D85481; D85481.
DR   PIR; D90630; D90630.
DR   GO; GO:0003773; F:heat shock protein activity; IEA.
KW   Heat shock; Complete proteome.
SQ   SEQUENCE   206 AA;  22547 MW;  C8C7BC70F7268EB1 CRC64;

Alignment Scores:
Pred. No.:                0.0637          Length:         206
Score:                    127.50          Matches:        63
Percent Similarity:       38.84%          Conservative:   24
Best Local Similarity:    28.12%          Mismatches:     96
Query Match:              9.46%           Indels:         41
DB:                       16              Gaps:           7

US-09-732-091-3 (1-759) x Q8XA70 (1-206)

Qy        629 ACGCCTGTAATGATCCAGCCAACAGGACCTGTTAAAAAGCTCAGAGTTCTTGTAAGCACC 570
              ||||||              ||||||||||||
Db          5 ThrProPheSerAlaAlaProThrGlyProProSer------------------------ 16

Qy        569 TGATTGCCCGCA---AGCGATAAACCACGCCCTAGAATGGTTTTTGCGACCGCATTCGCA 513
                  ||||||      ||||||   |||       :::   |||
Db         17 ------ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaTrpValArg 34

Qy        512 ACAATGACAGCTAATTGATAAGATTTAAAA------CCCCCCATTTTAAACAGCGTTAAA 459
                ||      :::     :::   :::          |||     |||:::
Db         35 LysMetArgValSerTrpLeuGluSerArgCysAspThrProPheAlaAsnAsn------ 52

Qy        458 GTCGCCGCGCTTAAGGCTTGTCTGTTTAAATTGTCCGTGTTTTTTATGGATAATTCATCG 399
                                          ||||||                 :::||||||
Db         53 ----------------------------LeuSerPheIleSerSerGlySerSerSer 62

Qy        398 CACATTTCTTTCACT--------------------------------TCTTCATCATCC 372
              |||||||||                                    |||||||||
Db         63 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuCysSerSerSer 82

Qy        371 ATTTCTTCCAAACTTCTTTCTAAGATTTTAGAAAGCATGTTTTGTTCAATTAAAGTCGTT 312
              |||      ::::::|||   :::         |||      |||||||||      |||
Db         83 IlePhePheGlnValLeuArgArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 102

Qy        311 TCAGTTTTCTTGTTGTAATTGACCTTTAATTTATCGCACACATCGCATAAAATCTCTTTG 252
              :::   |||         :::||||||      ||||||   |||::::::
Db        103 AspIleSerLeuSerAlaPheSerPheAsnArgPheGluThrSerSerLysMetAlaArg 122

Qy        251 TATAAGACTCCTTCGCCTTTAATGAAACTCGCAAAACTATTGCTCCCATAGTATTGCAAC 192
              |||      |||   |||         ||||||   |||      |||      ||||||
Db        123 TyrAsnLeuProCysProArgSerLeuLeuAlaIleLeuSerProProLysCysCysAsn 142

Qy        191 TCTTCAGCGATTCTTTCTGCGTATTTAGCGTAATCATCGCCATGC------------CTT 144
              |||   ||||||                        |||   |||            |||
Db        143 SerProAlaIleSerCysGlnLeuArgArgCysCysSerGlyCysProSerIleAspLeu 162

Qy        143 TTGTATTCTATGGAGCTGGTCAGTTTTTCATTGTGTCTTTTTTCGCCGTCTTTACCAAAA 84
                |||:::   :::   :::             :::         ||||||
```

```
Db        163 AsnSerSerLeuArgIleSerThrLeuGluArgArgValLeuProPheSerLeuTrpVal 182

Qy         83 ACAAGCACCTCAAACAAATCCAATAAATCACTAGATTCCAATTGCTTTAAAAATTCCAAG 24
              :::|||   :::     :::|||   |||            |||::::::   |||:::
Db        183 SerSerArgAlaLysPheAlaAsnCysSer---SerLeuGlnCysTrpArgLysSerArg 201

Qy         23 TCTCTATCATAT 12
              |||   |||:::
Db        202 SerGluSerPhe 205
```

RESULT 6
Q7UDU2
ID Q7UDU2 PRELIMINARY; PRT; 199 AA.
AC Q7UDU2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Positive regulator for sigma 32 heat shock promoters.
GN HTGA OR S0011.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AE016978; AAP15558.1; -.
KW Heat shock.
SQ SEQUENCE 199 AA; 21606 MW; 3598441A948498B4 CRC64;

Alignment Scores:

Pred. No.:	0.101	Length:	199
Score:	125.00	Matches:	63
Percent Similarity:	36.49%	Conservative:	18
Best Local Similarity:	28.38%	Mismatches:	81
Query Match:	9.27%	Indels:	60
DB:	16	Gaps:	8

US-09-732-091-3 (1-759) x Q7UDU2 (1-199)

```
Qy        629 ACGCCTGTAATGATCCAGCCAACAGGACCTGTTAAAAAGCTCAGAGTTCCTGTAAGCACC 570
              |||||              ||||||||||
Db          8 ThrProPheAsnAlaAlaProThrGlyProProSer---------------------- 19

Qy        569 TGATTGCCCGCA---AGCGATAAACCACGCCCTAGAATGGTTTTTGCGACCGCATTCGCA 513
                  ||||||          ||||||   |||      :::   |||
Db         20 ------ProAlaProArgSerLysProCysProSerThrLeuIleAlaAlaTrpValArg 37

Qy        512 ACAATGACAGCTAATTGATAAGATTTAAAA------CCCCCCATTTTAAACAGCGTTAAA 459
                 |||      :::      :::   |||          |||      |||:::
Db         38 LysMetArgValSerTrpLeuGluSerLysCysAspThrProPheAlaAsnAsn------ 55

Qy        458 GTCGCCGCGCTTAAGGCTTGTCTGTTTAAATTGTCCGTGTTTTTTATGGATAATTCATCG 399
                                                 ||||||               :::||||||
Db         56 ----------------------------LeuSerPheIleSerSerGlySerSerSer 65

Qy        398 CACATTTCTTTCACT------------------------------TCTTCATCATCC 372
              |||||||||                                    |||||||||
Db         66 SerSerSerPheThrLeuAlaSerThrAlaCysArgAsnSerCysLeuCysSerSerSer 85

Qy        371 ATTTCTTCCAAACTTCTTTCTAAGATTTTAGAAAGCATGTTTTGTTCAATTAAAGTCGTT 312
                 |||      ::::::|||   :::         |||      |||||||||      |||
Db         86 IlePhePheGlnValLeuArgArgAsnCysSerSerAsnCysCysSerIleSerAsnVal 105

Qy        311 TCAGTTTTCTTGTTGTAATTGACCTTTAATTTATCGCACACATCGCATAAAATCTCTTTG 252
```

```
                :::   |||         :::||||||        ||||||   |||::::::
Db        106 AspIleSerLeuSerAlaPheSerPheAsnArgPheGluThrSerSerLysMetAlaArg 125

Qy        251 TATAAGACTCCTTCGCCTTTAATGAAACTCGCAAAACTATTGCTCCCATAGTATTGCAAC 192
              |||      |||  |||         ||||||  |||       |||      ||||||
Db        126 TyrAsnLeuProCysProArgSerLeuLeuAlaIleLeuSerProProLysCysCysAsn 145

Qy        191 TCTTCAGCGATTCTTTCTGCGTATTTAGCGTAATCATCGCCATGC--------------- 147
              |||   ||||||                        |||   |||
Db        146 SerProAlaIleSerCysGlnLeuArgArgCysCysSerGlyCysProSerIleAspLeu 165

Qy        146 ---------CTTTTGTATTCTATG---------GAGCTGGTCAGTTTTTCATTGTGTCTT 105
                      |||      ||||||             ::::::   |||||||||   :::
Db        166 AsnSerSerLeuArgIleSerMetLeuGluArgArgValLeuProPheSerLeuTrpVal 185

Qy        104 TTTTCGCCGTCTTTACCAAAAACAAGCACCTCAAACAAATCCAATAAATCACTAGATTCC 45
                                                  ||||||::::::      :::
Db        186 ----------------------------------------SerAsnArgAlaLysPheAla 192

Qy         44 AATTGC 39
              ||||||
Db        193 AsnCys 194
```

RESULT 7
Q8IAN1
ID Q8IAN1 PRELIMINARY; PRT; 1245 AA.
AC Q8IAN1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF08_0127.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL844507; CAD51332.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1245 AA; 147911 MW; D856486AFDFE4DDF CRC64;

Alignment Scores:

Pred. No.:	3.84	Length:	1245
Score:	105.00	Matches:	35
Percent Similarity:	42.42%	Conservative:	21
Best Local Similarity:	26.52%	Mismatches:	48
Query Match:	7.78%	Indels:	28
DB:	5	Gaps:	5

US-09-732-091-3 (1-759) x Q8IAN1 (1-1245)

```
Qy        100 GAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGAT--- 156
              ||||||:::      ||||||:::          :::   |||      |||   |||
Db        542 GluLysLysGluGluGluLysIleMetLysLysLeu---TyrSerGluHisProAspLeu 560

Qy        157 ------GATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGC 210
                    |||      |||   :::      ||||||:::                 :::
Db        561 ValLysAspLeuMetAsnTyrGluLysGluTyrAlaGluLysArgAsnLeuLeuAsnAsn 580

Qy        211 AATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGAT 270
                          ||||||               ||||||
Db        581 IleLysLysArgLysThrIleLysThrLeuAsn---ArgTyrLysAsn------------ 595

Qy        271 GTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATTGAACAA 330
                    ::::::||||||            |||      ::::::         |||:::
Db        596 ------GluGlnLeuLysLysGluIleLeuLysSerIleGlnSerAsnLysGluGluLys 613

Qy        331 AACATGCTTTCTAAAATCTTAGAAAGAAGTTTG-------------------------- 363
              |||||||||   ||||||:::|||::::::
Db        614 AsnMetLeuLysLysIleIleGluLysThrAlaLeuThrGluGlnThrAsnGluAspAsn 633
```

```
Qy        364 ----------------------GAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGC 399
                                     ::::::  ||||||:::  :::  |||
Db        634 AsnMetMetIleHisAspAspAsnAspAspAspAspAspAspAsnMetSerGluAspThr 653

Qy        400 GATGAATTATCCATAAAAAACACGGACAATTTAAAC 435
              ||||||:::|||  |||:::  |||||:::|||
Db        654 AspGluIleSerSerLysAspHisAspAsnMetAsn 665
```

RESULT 8
Q8R9Z4
ID Q8R9Z4 PRELIMINARY; PRT; 403 AA.
AC Q8R9Z4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein TTE1436.
GN TTE1436.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013102; AAM24658.1; -.
DR GO; GO:0009424; C:flagellar hook (sensu Bacteria); IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0009296; P:flagella biogenesis; IEA.
DR InterPro; IPR001635; Flag_hook.
DR Pfam; PF02120; Flg_hook; 1.
DR PRINTS; PR01007; FLGHOOKFLIK.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 403 AA; 46116 MW; 60C22B9FBB51A56F CRC64;

Alignment Scores:

Pred. No.:	5.08	Length:	403
Score:	103.50	Matches:	63
Percent Similarity:	35.14%	Conservative:	47
Best Local Similarity:	20.13%	Mismatches:	98
Query Match:	7.67%	Indels:	105
DB:	16	Gaps:	12

US-09-732-091-3 (1-759) x Q8R9Z4 (1-403)

```
Qy         28 GAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTTTTT 87
              |||||||||  |||                  :::::::::|||:::
Db         61 GluPheLeuSerGlnLysAsnGlyLeuLysAsnValGluIlePheGln------------ 76

Qy         88 GGTAAAGACGGCGAAAAAAGACACAATGAAAAA-------------------------- 120
                      ||||||         ::|||
Db         77 ------------GluLysValLysGluAspLysAspIleMetGluAspLeuAsnAsnPhe 92

Qy        121 ---------------------CTGACCAGCTCCATAGAA---------TACAAAAGGCAT 150
                                  :::|||  |||:::|||           ::::::
Db         93 IleProAlaLeuMetGlnThrIleThrLeuSerLeuGluThrIleSerSerGluLysGly 112

Qy        151 GGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGC 210
                 |||:::  |||      |||::::::       ||||||
Db        113 AlaSerAspPheGluLysValArgGluLysLeuGluValAlaLeuGln------------ 128

Qy        211 AATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGAT 270
                       ||||||||           :::||||||||
Db        129 ------------GlyPheIleLysGluArgAsnPheThrPheLysGluIleAlaLysLys 144

Qy        271 GTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATTGAACAA 330
              :::  |||  ||||||  |||:::|||  :::           :::||||||:::
Db        145 IleSerAspPheLeuLysGluAsnPheAsnIleGluLeuSerProGluValIleGluArg 164

Qy        331 AACATG---CTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTG 387
              ::::::   |||:::|||:::  :::          :::  |||
Db        165 HisIleLysLeuAlaLysValLysAspLeuAspLysProPheLeuGlnAspLeuAsnGln 184

Qy        388 AAAGAAATGTGCGATGAATTATCCATAAAAAACACGGAC--------------AATTTA 432
              |||:::      :::|||         ||||||  |||                :::
Db        185 LysAspPheAlaGluGluAsnValGlnLysAsnGlnAspLysThrSerGlnLeuLysIle 204

Qy        433 AACAGACAAGCCTTAAGCGCGGCG---------------------------------- 456
              :::::::::|||      ||||||
Db        205 AspLysGluAlaPheIleAlaAlaLysGluAlaLysGluGluLysThrGluLysLysSer 224

Qy        457 ---------------ACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTA 501
                            ||||||      |||                :::
Db        225 PheAspValLysGlnGluPheValPhePheLysAsnGluGly---------LysProVal 241

Qy        502 GCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCG 561
              :::   :::     |||::::::   |||:::
Db        242 SerAsnLeuThrTyrAsnSerIleLysLysSer---------------------------- 252

Qy        562 GGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATT 621
              ||:      :::  |||                                    ||:::
Db        253 ---AsnAspProValAspArgLeuPheArg--------------------GlnIleVal 264

Qy        622 ACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCA 681
                  |||:::  |||  :::          :::      |||  :::
Db        265 AspAsnValPheValAlaLysGluLysGlyAlaSerSerValThrValAsnLeuLysPro 284

Qy        682 TGCATT-----------------------------------------GTGGTTGCCACT 699
                  ||                                       :::||||||||||
Db        285 GluIleLeuGlyLysLeuGlnIleSerLeuLysSerIleAspGlyAsnIleValAlaThr 304

Qy        700 TTA-----------------------------------CGCCTAAAAACACAGCAA 720
              :::                                  |||         |||
Db        305 IleValThrGluSerGluLysThrLysHisGlnIleGluSerAsnLeuSerLeuLeuGln 324

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              |||      |||  |||  ::::::|||||||||:::
Db        325 AlaGlnLeuAspLeuLysGlyIleLysIleGluSerVal 337
```

RESULT 9
Q8DIH4
ID Q8DIH4 PRELIMINARY; PRT; 276 AA.
AC Q8DIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Tll1614 protein.
GN TLL1614.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005374; BAC09166.1; -.
KW Complete proteome.
SQ SEQUENCE 276 AA; 31282 MW; D0577A9D97E0CA92 CRC64;

Alignment Scores:

Pred. No.:	6.11	Length:	276
Score:	102.50	Matches:	50
Percent Similarity:	39.16%	Conservative:	53

```
Best Local Similarity:  19.01%          Mismatches:     111
Query Match:            7.60%           Indels:         49
DB:                     16              Gaps:           10

US-09-732-091-3 (1-759) x Q8DIH4 (1-276)

Qy          34 TTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAA 93
               |||:::      :::|||   ||||||   |||:::|||
Db          24 LeuGluLeuAlaThrGluGluGluLeuGlnAspLeuThrGluIleLeuPhe--------- 40

Qy          94 GACGGCGAAAAAAGACACAAT------GAAAAACTGACCAGCTCCATAGAATACAAAAGG 147
                  :::|||   |||      :::   ||||||:::
Db          41 ------ArgArgArgLeuAsnProLeuAspTyrLeuThrThrProAspProIleAlaVal 58

Qy         148 CATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGG 207
                  |||         :::   :::   |||   |||       ::::::
Db          59 GlnAlaGlnAspArgGlnAlaTrpLeuAspAspIleGluGluArgPheArgPheLeuAla 78

Qy         208 AGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGA------GTCTTATACAAAGAGATT 261
               ::::::      :::   :::|||||:::        :::   |||::::::
Db          79 AlaAspGlyLeu---ThrValLeuLysGlyLysAlaGlnGlnIleSerTyrArgGlnThr 97

Qy         262 TTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTA 321
               |||      ||||||      ||||||:::   ::::::
Db          98 LeuMetArgValCysArgTyrLeuLysIleLysPheSerProSerTrpThrValProGlu 117

Qy         322 ATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAA 381
               :::|||      :::         :::|||:::|||      :::::::::   |||:::
Db         118 LeuGluMetGluIlePheLeuAsnValLeuGlnArgMetTrpLysLysLeuGlyAspGln 137

Qy         382 GAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAACACGGACAATTTAAACAGACAA 441
               :::   :::   :::   ::::::         :::   :::
Db         138 AspArgArgValLeuAlaAlaGlnIleGlnGluSerLeuProGluLeuHisHisGlyHis 157

Qy         442 GCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTA 501
              :::|||         :::   |||         ||||||
Db         158 ProIleSerMetGluMetValArgLeuValLeuGluGlyGly------------------ 171

Qy         502 GCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA--------------------- 540
               |||   |||      ::::::|||   ::::::::::::
Db         172 AlaAlaIleAlaIleSerSerValValArgSerMetValValGlnGlnValAlaArgGln 191

Qy         541 ------------------GGGCGTGGTTTATCGCTTGCG---------------------- 561
                                  |||      ||||||:::|||
Db         192 PheAlaIleArgPheAlaGlySerLysLeuSerIleAlaProLeuValSerArgGlyAla 211

Qy         562 ------------GGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTT 609
                           :::   |||::::::
Db         212 AlaMetGlyValAlaArgLeuAlaValGlyArgSerIleLeuAlaPheValSerThrAla 231

Qy         610 GGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT--------- 660
               |||              ||||||         |||:::        |||
Db         232 LeuTrp------------ValTrpPheIleAlaAspLeuGlyTrpGlnAlaIleSerThr 247

Qy         661 ---TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGC---CTAAAAACA 714
                  |||         |||||         :::|||   :::|||   |||:::
Db         248 AsnTyrAlaArgIleIleProThrIlePheAlaIleAlaGlnIleArgLeuLeuArgGly 267

Qy         715 CAGCAAGCC 723
               :::||||||
Db         268 GluGlnAla 270

RESULT 10
Q8MX30
ID   Q8MX30      PRELIMINARY;      PRT;  1590 AA.
AC   Q8MX30;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RNA polymerase II largest subunit (Fragment).
GN   RPB1.
CS   Naegleria gruberi.
OC   Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.
OX   NCBI_TaxID=5762;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12032239;
RA   Dacks J.B., Marinets A., Ford Doolittle W., Cavalier-Smith T.,
RA   Logsdon J.M. Jr.;
RT   "Analyses of RNA Polymerase II Genes from Free-Living Protists:
RT   Phylogeny, Long Branch Attraction, and the Eukaryotic Big Bang.";
RL   Mol. Biol. Evol. 19:830-840(2002).
DR   EMBL; AF395110; AAM45151.1; -.
DR   GO; GO:0005665; C:DNA-directed RNA polymerase II, core complex; IEA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0006366; P:transcription from Pol II promoter; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   InterPro; IPR006592; RNA_polA_N.
DR   InterPro; IPR000684; RNA_polII_repeat.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR007080; RNA_pol_Rpb1_1.
DR   InterPro; IPR007066; RNA_pol_Rpb1_3.
DR   InterPro; IPR007083; RNA_pol_Rpb1_4.
DR   InterPro; IPR007081; RNA_pol_Rpb1_5.
DR   InterPro; IPR007075; RNA_pol_Rpb1_6.
DR   InterPro; IPR007073; RNA_pol_Rpb1_7.
DR   Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR   Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR   Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR   Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR   Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR   Pfam; PF04992; RNA_pol_Rpb1_6; 1.
DR   Pfam; PF04990; RNA_pol_Rpb1_7; 1.
DR   Pfam; PF05001; RNA_pol_Rpb1_R; 5.
DR   SMART; SM00663; RPOLA_N; 1.
DR   PROSITE; PS00115; RNA_POL_II_REPEAT; 7.
FT   NON_TER       1      1
FT   NON_TER    1590   1590
SQ   SEQUENCE   1590 AA;  177446 MW;  5F4C7959D4D77D68 CRC64;

Alignment Scores:
Pred. No.:              6.64            Length:         1590
Score:                  102.00          Matches:        60
Percent Similarity:     35.69%          Conservative:   36
Best Local Similarity:  22.30%          Mismatches:     71
Query Match:            7.56%           Indels:         102
DB:                     5               Gaps:           13

US-09-732-091-3 (1-759) x Q8MX30 (1-1590)

Qy          82 GTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATAC 141
               ::::::|||:::||||||         |||       :::   :::   ||||||:::
Db         810 LeuTyrGlyGluAspGlyMetAspAlaHisLysValGlu---SerGlnSerIleAspIle 828

Qy         142 AAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATAC 201
                     |||      |||||:::   ::::::
Db         829 MetAsnLeuSerAsp------AlaLysPheGlnAspLys---------------------- 839

Qy         202 TATGGGAGCAATAGTTTTGCGAGT---------------------------------- 225
               |||      ||||||     |||
Db         840 TyrTrpTyrAsnSerLeuGlySerSerProSerPheThrAsnProThrGluSerSerVal 859

Qy         226 TTCATTAAAGGCGAAGGAGTCTTA--------------------------TACAAAGAG 258
                  :::         |||:::|||                        |||   |||
Db         860 SerMetProSerGlyGlyIleLeuLysGluSerPheIleSerThrLysValTyrAspGlu 879

Qy         259 ATTTTATGCGATGTGTGCGATAAATTA---AAGGTCAATTACAACAAGAAAACT------ 309
               ||||||   |||            |||   ||||||          ||||||
Db         880 IleLeuLysAspProAlaThrTyrLeuSerLysValArgGluGluTyrLysThrLeuMet 899

Qy         310 ----------------------------------------GAAACGACTTTAATTGAA 327
                                                      |||      ::::::
Db         900 GluAspArgHisIleLeuArgThrGluIleProAsnAlaGluAsnLysIleValMet 919
```

```
Qy        328 CAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATGGAT---------- 375
                  |||   :::|||::::::   :::   :::|||
Db        920 ProValAsnLeuLysArgIleIleLysAsnAlaGlnLysGluTyrGlyIleHisProSer 939

Qy        376 ----------------------------------GATGAAGAAGTGAAAGAAATGTGC 399
                                                |||:::::|||:::::||
Db        940 IleGlyLysProSerAspMetAsnProIleTyrValIleGluLysIleLysLysLeuCys 959

Qy        400 GATGAATTA---TCCATAAAAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCG 456
                |||:::|||      ||||||  ||||||  |||:::  :::|||  :::   |||
Db        960 AspAspLeuValValIleLysGlyThrAspGluLeuSerThrGluAlaGlnThrAsnAla 979

Qy        457 ACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCG 516
                ||||||  ||||||  |||
Db        980 ThrLeu---LeuPheSerMet---------------------------------------- 985

Qy        517 AATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTT 576
                  :::     |||       :::|||  ||||||          :::|||
Db        986 ---PheLeuArgSerThrPhePhePheLysGlyCysSerLeuLysAsnCysValLeuLeu 1004

Qy        577 ACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATT--------------- 621
                  :::   |||:::|||             |||::::::
Db       1005 LysLysProLeuThrPhe-----------------TrpLeuValLysLeuGluLeuArg 1018

Qy        622 --------------------------------ACAGGCGTATGGACAGCGATTGATATT 648
                                              |||     |||     |||
Db       1019 PheHisAsnCysLeuAlaGlnProGlyGluMetValGlySerValAlaAlaGlnSerIle 1038

Qy        649 GCAGGGCCGGCTTATAGGGTAACCATA 675
                  ||||||   ::::::|||:::
Db       1039 GlyGluProAlaThrGlnMetThrLeu 1047

RESULT 11
Q8IKG8
ID   Q8IKG8      PRELIMINARY;      PRT;  1455 AA.
AC   Q8IKG8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Rhoptry protein, putative.
GN   PF14_0637.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014826; AAN37250.1; -.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.
DR   InterPro; IPR002114; HPr_SerP_S.
DR   PROSITE; PS00589; PTS_HPR_SER; 1.
SQ   SEQUENCE   1455 AA;  174305 MW;  36FAA1760A898F99 CRC64;

Alignment Scores:
Pred. No.:              7.97          Length:        1455
Score:                  101.00        Matches:       36
Percent Similarity:     46.10%        Conservative:  29
Best Local Similarity:  25.53%        Mismatches:    52
Query Match:            7.49%         Indels:        24
```

```
DB:                     5             Gaps:          6

US-09-732-091-3 (1-759) x Q8IKG8 (1-1455)

Qy         16 GATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAG 75
              :::|||:::|||   :::|||      :::|||               :::|||
Db        333 GluArgGluLeuTyrTyrLeuGlyMetIleGlu--------------GluLeuArgAsn 347

Qy         76 GTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATA 135
                 :::         ::::::   |||      :::   |||||||||   :::   |||
Db        348 GluIleLysThrLysGluGluAsnGluGlyAsnAsnIleGluLysLeuGluAsnLysIle 367

Qy        136 ---GAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAG 192
                  |||||::::::      ::::::              |||::::::
Db        368 HisGluTyrGluLysGlnAsnGluGluLeuArgAsnGluLysGluLysLeuGlnSerThr 387

Qy        193 TTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATAC 252
                 :::      |||   |||::::::|||  :::                  :::
Db        388 IleAsnGluTyr---SerHisAsnPheAsnAsnLeuAsnAspHisAsnLysIleThrAsn 406

Qy        253 AAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC---------AAG 303
                 ||||||                |||::::::||||||   |||||||||          :::
Db        407 LysGlu---------------CysGluGluLeuLysAsnAsnTyrAsnThrIleLysGlu 421

Qy        304 AAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTG 363
                 |||   |||                |||||   :::     |||
Db        422 LysTyrGluArgLeuLysGluGluGlnGluIleTyrIleLys------------------ 435

Qy        364 GAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACG 423
                          :::||||||   |||   :::   |||||||||      :::|||
Db        436 ---------GlnGluGluGluTyrLysSerLeuLeuAspGluLeuGluAsnGluAsnAsn 452

Qy        424 GAC 426
                 :::
Db        453 Glu 453

RESULT 12
P87397
ID   P87397      PRELIMINARY;      PRT;   726 AA.
AC   P87397;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Heat shock protein hsp90.
OS   Oncorhynchus tschawytscha (Chinook salmon) (King salmon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=74940;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99262176; PubMed=10329464;
RA   Palmisano A.N., Winton J.R., Dickhoff W.W.;
RT   "Sequence features and phylogenetic analysis of the stress protein
RT   hsp90alpha in chinook salmon (Oncorhynchus tshawytscha), a
RT   poikilothermic vertebrate.";
RL   Biochem. Biophys. Res. Commun. 258:784-791(1999).
DR   EMBL; U89945; AAB49983.1; -.
DR   HSSP; P07900; 1BYQ.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003754; F:chaperone activity; IEA.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR001404; Hsp90.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00183; HSP90; 1.
DR   PRINTS; PR00775; HEATSHOCK90.
DR   SMART; SM00387; HATPase_c; 1.
DR   PROSITE; PS00298; HSP90; 1.
SQ   SEQUENCE   726 AA;  83505 MW;  1A2A962C471CBCF2 CRC64;

Alignment Scores:
Pred. No.:              8.76          Length:        726
Score:                  100.50        Matches:       61
```

```
Percent Similarity:     35.46%          Conservative:  39
Best Local Similarity:  21.63%          Mismatches:    81
Query Match:            7.45%           Indels:        101
DB:                     13              Gaps:          11

US-09-732-091-3 (1-759) x P87397 (1-726)

Qy          61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAAGACACAATGAAAAA 120
               :::|||||||||       |||         ||||||       ||:::      |||:::
Db         414 MetAspLeuPheValGluLeuSerGluAspLysAspAsnTyrLysLysPheTyrGluGln 433

Qy         121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
                  :::   :::|||:::         |||   |||      |||:::      ::::::
Db         434 PheSerLysAsnIleLysLeuGlyIleHisGluAsp------AlaGlnAsnArgLysLys 451

Qy         181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
               :::::::::   |||:::||||||   |||||||||   |||                |||
Db         452 LeuSerAspMetLeuArgTyrTyrThrSerAsnSerAsnAlaAsp-----------Glu 467

Qy         241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
                  |||       ||||||   :::       :::   |||       |||       |||
Db         468 MetValSerLeuLysGluTyrValSerArgMetLysAspThrGlnLysHisIleTyrTyr 487

Qy         301 AAGAAAACTGAAACGACT---------------------------------------- 318
                     ||||||
Db         488 IleThrGlyGluThrLysGluGlnValAlaAsnSerSerPheValGluArgLeuArgLys 507

Qy         319 ----------------------TTAATTGAA---------------------------- 327
                                    :::||||||
Db         508 AlaGlyLeuGluValIleTyrMetIleGluProIleAspGluTyrCysValGlnGlnLeu 527

Qy         328 ---------------CAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATG 372
                           :::|||::::::|||   :::   :::      ||||||
Db         528 LysGluTyrAspGlyLysAsnLeuValSer---ValThrLysGluGlyLeuGluLeuPro 546

Qy         373 GATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACGGACAATTTA 432
               :::||||||:::   |||:::      :::||||||:::   |||      :::||||||
Db         547 GluAspGluAspGluLysLysLysGlnGluGluLeuAsnThrLys---PheGluAsnLeu 565

Qy         433 AACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCT 492
               :::
Db         566 CysLys---------------------------------------------------- 567

Qy         493 TATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTA 552
                                                 |||   ||||||   :::   :::
Db         568 ------------------------------ThrMetLysAspIleLeuAspLysLysIle 577

Qy         553 TCG------CTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCT 606
                        |||:::::::::::                :::|||
Db         578 GluLysValSerValSerAsnArgLeuValSerSerProCysCysIleValThrSerThr 597

Qy         607 GTTGGCTGGATCATTACA------------------------------------------ 624
               ||||||
Db         598 TyrGlyTrpThrAlaAsnMetGluArgIleMetLysSerGlnAlaLeuArgAspAsnSer 617

Qy         625 ------GGCGTATGGACAGCG---------ATTGATATTGCAGGGCCGGCTTATAGGGTA 669
                    |||      ||||||         ::::::|||
Db         618 ThrMetGlyTyrMetThrAlaLysLysHisLeuGluIleAsn----------------- 631

Qy         670 ACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAAGCCAATGGA 729
                     |||         :::|||   |||||||||   |||   :::
Db         632 ------ProThrHisProIleValGluThrLeuArgGluLysAlaGluAlaAspLysAsn 649

Qy         730 GATAAG 735
               ||||||
Db         650 AspLys 651

RESULT 13
Q7VHM0
ID   Q7VHM0      PRELIMINARY;      PRT;   556 AA.
AC   Q7VHM0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   HH0946.
OS   Helicobacter hepaticus.
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=32025;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 51449 / 3B1;
RX   MEDLINE=22709201; PubMed=12810954;
RA   Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA   Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA   Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA   Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT   "The complete genome sequence of the carcinogenic bacterium
RT   Helicobacter hepaticus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR   EMBL; AE017147; AAP77543.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   556 AA;  61984 MW;  0E48CF5376BAC48D CRC64;

Alignment Scores:
Pred. No.:              10.5            Length:        556
Score:                  99.50           Matches:       53
Percent Similarity:     38.57%          Conservative:  33
Best Local Similarity:  23.77%          Mismatches:    104
Query Match:            7.38%           Indels:        33
DB:                     16              Gaps:          6

US-09-732-091-3 (1-759) x Q7VHM0 (1-556)

Qy          94 GACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGC 153
               |||      |||||||||   |||   |||         ||||||      ::::::
Db         199 AspAspTyrLysArgHisLysGluAsnLeuGluLysMetIleGluLysGlyLysAsnAsn 218

Qy         154 GATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAAT 213
                  :::                                                   |||
Db         219 ProGlu-----------------------------------------------AsnGlnAsn 223

Qy         214 AGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTG 273
               :::         |||   |||   |||   |||                  ::::::
Db         224 AlaValGluLysAlaGlnLysAlaGluLysAlaLeuGluLysLeuAsnAlaAsnAsnMet 243

Qy         274 TGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTA------------ 321
               |||::::::|||                :::   |||
Db         244 CysAsnArgLeuMetCysGluAsnProArgIleThrAlaValAlaThrGlnSerIleVal 263

Qy         322 ------------ATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAA 369
                           |||   |||         |||      :::      :::|||
Db         264 AlaSerGlyHisIleAlaGlnAlaGlyPheSerAspAlaIleIleAlaAlaLeuAlaThr 283

Qy         370 ATGGATGATGAA---------GAAGTGAAAGAAATGTGCGAT---------------GAA 405
               :::   :::            |||||||||:::|||   :::               |||
Db         284 LeuAlaAsnGlyValIleTrpGluValLysAspMetPheGluGlySerIleAspThrGlu 303

Qy         406 TTATCCATA---AAAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTA 462
                  ||||||   |||         |||   ::::::   :::   |||
Db         304 ThrSerIleLeuLysArgIleLysArgLeuLeuLysArgThrIleGluAlaPheGlnAla 323

Qy         463 ACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCG 522
               |||      :::      ||||||   :::         ||||||   |||
Db         324 ThrPheGlyArgGlyAlaGlyPheGlyAlaIleAspAlaAlaValGlyValValGlyGln 343

Qy         523 GTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGA 582
               :::   ::::::|||   |||:::   |||   |||      :::::::::   |||
Db         344 IlePheArgSerIleAlaGlyLys---LeuLysLeuValTrpAspLysIleArgThrAla 362

Qy         583 ACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATT 642
                  |||         |||   |||   :::|||         |||
Db         363 LeuLysSerIleTyrAsnGlyIleValSerTyrIleLysGlyGluValSerAsnLeuArg 382

Qy         643 GATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTA 702
```

```
Db        383 GluLeuLeuGlyIleIleLeuLysSerLeuPheSerAlaAlaTrpValValSerThrLeu 402

Qy        703 CGCCTAAAA 711
                 |||:::
Db        403 AlaLeuGlu 405
```

RESULT 14
Q882L0
ID Q882L0 PRELIMINARY; PRT; 633 AA.
AC Q882L0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN PSPTO2616.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White C., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE016865; AAO56120.1; -.
DR TIGR; PSPTO2616; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 1.
KW Complete proteome.
SQ SEQUENCE 633 AA; 68090 MW; D15A8BC00AC58637 CRC64;

Alignment Scores:

Pred. No.:	10.5	Length:	633
Score:	99.50	Matches:	69
Percent Similarity:	40.89%	Conservative:	50
Best Local Similarity:	23.71%	Mismatches:	87
Query Match:	7.38%	Indels:	85
DB:	16	Gaps:	18

US-09-732-091-3 (1-759) x Q882L0 (1-633)

```
Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
              |||     :::|||     ::::::||||||   ||||||     :::
Db         70 TyrAspAlaGluArgAlaSerAspTrpLeuLysAlaLeuGluArgLeuGluAlaHisVal 89

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
                 :::       ||||||           :::         |||      |||
Db         90 LysTyrAlaGlnGluValPhe----------AspSerProLeuAsnIleProLeuVal 105

Qy        127 AGCTCC---------------ATAGAATACAAAAGGCATGGCGATGAT------------ 159
              ::::::                :::|||||:::   |||   |||:::
Db        106 AsnAlaAlaAlaGluAlaLeuValGluTyrArgValHisTyrAspAsnLeuMetArgAla 125

Qy        160 ----------------------------TACGCTAAATACGCAGAAAGAATCGCTGAA 189
                                         :::   :::||||||:::      ||||||
Db        126 ThrAlaAlaArgGluAlaThrArgGlyAlaPheGlyGlnTyrAlaAspAlaGlyAlaGlu 145

Qy        190 GAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGA------ 243
              :::||||||           |||:::||||||        :::|||
```

```
Db        146 AspLeuGlnLysLeu------AsnAlaPheAlaArgSerAspGluGlySerAlaSerGln 163

Qy        244 ----------------------GTCTTATACAAAGAGATTTTATGCGATGTGTGCGAT 279
                                     |||:::::::::::      |||:::
Db        164 ArgAspAlaIleValGlnAlaMetThrLeuPheGlnLysMetArgPheAspLeu------ 181

Qy        280 AAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTT 339
                        |||:::   |||   |||                :::       :::
Db        182 ---ArgGlyTyrThrTyrSerLeuLysAlaGluAsnArgAlaProAlaGluAlaSerMet 200

Qy        340 TCTAAAATCTTAGAA------AGAAGTTTGGAAGAAATGGATGATGAAGAA------GTG 387
              |||   ::::::         :::   |||:::      |||   :::         :::
Db        201 SerAlaValIleAsnPheValLysGlyLeuGlnGlyPheAspSerGlnSerAlaThrIle 220

Qy        388 AAAGAAATGTGCGATGAATTA---TCCATAAAAAACACGGACAAT---------TTAAAC 435
              |||   :::   |||   :::   |||   :::||||||   |||
Db        221 LysHisLeuValAspSerMetValSerTyrGlnAsnThrMetAsnGlnPheThrAlaAla 240

Qy        436 AGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGT------TTTAAA 489
              :::   ::::::       |||       :::   |||:::   |||       ||||||
Db        241 GlnAlaSerIleAspGlnAlaGlnAlaGlyIleThrLysValIleGlyValLeuPheLys 260

Qy        490 TCTTAT---CAATTAGCTGTCATTGTTGCGAATGCGGTC---------------------- 525
              |||       ||||||:::         |||||   |||
Db        261 SerAlaAspGlnLeuSer---------AlaAsnGlnValSerLeuArgIleGluAspVal 277

Qy        526 ------GCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACA 579
                    |||:::|||:::|||         |||                   |||
Db        278 AspGlnAlaArgThrLeuLeuSerValTrpLeu-----------------ValAlaAla 291

Qy        580 AGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCG 639
                  :::|||||||||         |||:::||||||   :::
Db        292 LeuIleMetSerPheLeu---------AlaAlaTrpValIleThrArgLeu--------- 305

Qy        640 ATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACT 699
                  |||   ||||||                                ::::::   |||
Db        306 ------IleValGlyPro----------------------------LeuLeuGluThr 313

Qy        700 TTACGCCTAAAAACACAGCAAGCCAATGGAGAT 732
              |||:::|||       :::   |||:::||||||
Db        314 LeuLysLeuAlaGluArgValAlaAspGlyAsp 324
```

RESULT 15
Q8IJZ2
ID Q8IJZ2 PRELIMINARY; PRT; 202 AA.
AC Q8IJZ2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF10_0048.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014829; AAN35246.1; -.
KW Hypothetical protein.
SQ SEQUENCE 202 AA; 24259 MW; 48FD22A7F21165A6 CRC64;

Alignment Scores:

Pred. No.:	11.6	Length:	202
Score:	99.00	Matches:	46
Percent Similarity:	41.62%	Conservative:	31
Best Local Similarity:	24.86%	Mismatches:	50
Query Match:	7.34%	Indels:	58
DB:	5	Gaps:	9

US-09-732-091-3 (1-759) x Q8IUZ2 (1-202)

```
Qy         10 AAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTG 69
              ||||||:::  ::::::     ||||||:::|||      ::::::  :::
Db         32 LysTyrGluGluAsnIleIleLeuLeuLysLysLeuMetArgThrGluAsnIleLysPhe 51

Qy         70 TTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAAT--------------- 114
                                          ::::::::::::|||
Db         52 -----------------------------AspGlnLysTyrAsnAlaIleAsnIleLeu 61

Qy        115 GAAAAACTG------ACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAA 168
              |||||||||      |||:::|||:::         :::||||||||||||
Db         62 GluLysLeuLysAspIleGluAspIleGlnTyrGluAspIleLeuAsnAspTyrAlaLys 81

Qy        169 ------------------------TACGCAGAAAGAATCGCTGAAGAGTTGCAATACTAT 204
                                       |||::::::|||||||||:::|||
Db         82 AsnPheLeuThrPheLeuLysGlySerPheGluLysValAlaGluGluIleGlnIle--- 100

Qy        205 GGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTA 264
                 |||         ||||||   :::   |||      |||   |||:::   |||
Db        101 ------AsnAspLysIleSerPheArgArgAsnGlu------LeuIleLysGlnAspLeu 116

Qy        265 TGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAA------------------ 306
                :::           |||      |||||||||   :::
Db        117 AsnAsnLysLeuGlnAsnLeuAsnHisAsnTyrAsnLeuGluTyrIleAsnLeuAsnLys 136

Qy        307 ---------ACTGAA-----------------ACGACTTTAATTGAACAAAACATGCTT 339
                     ||||||                  ::::::   |||   |||
Db        137 CysLeuGlyThrGluAspPheGluSerGlnLeuAsnAlaSerLeuLysPheAsnSerLeu 156

Qy        340 TCTAAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAA--------- 390
                ||||||   :::   |||:::||||||   |||   |||::::::
Db        157 LysAspIleLeuSerLysLeuLeuLysGluMetLysAspTyrGluIleGlnTyrAsnAsn 176

Qy        391 ------------------------------GAAATGTGCGATGAATTATCCATAAAAAAC 420
                                          :::::|||:::      :::      |||
Db        177 LeuSerAspThrAspLysPheLeuAspMetLysIleCysGluLeuGluAsnTyrLeuAsn 196

Qy        421 ACGGACAATTTAAAC 435
              |||   |||
Db        197 AsnThrAsnGlnAsn 201
```

Search completed: July 5, 2004, 03:06:06
Job time : 87.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 5, 2004, 02:53:47 ; Search time 16.5 Seconds
(without alignments)
4749.589 Million cell updates/sec

Title: US-09-732-091-3
Perfect score: 1349
Sequence: 1 atggcatacaaatatgatag.....cgttgcaaatagaatccatt 759

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09732091/runat_02072004_182407_19046/app_query.fasta_1.903
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09732091_@CGN_1_1_18_@runat_02072004_182407_19046 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	364	27.0	273	4	US-09-252-991A-29841	Sequence 29841, A
2	346.5	25.7	321	4	US-09-252-991A-29840	Sequence 29840, A
3	247	18.3	258	4	US-09-489-039A-8852	Sequence 8852, Ap
4	103	7.6	739	4	US-09-543-681A-6437	Sequence 6437, Ap
5	101.5	7.5	173	2	US-08-658-639-12	Sequence 12, Appl
6	101.5	7.5	173	3	US-08-944-604-12	Sequence 12, Appl
7	99	7.3	222	3	US-08-944-604-16	Sequence 16, Appl
8	99	7.3	274	4	US-09-134-001C-5279	Sequence 5279, Ap
9	95	7.0	2125	4	US-09-919-172-29	Sequence 29, Appl
10	91.5	6.8	1086	4	US-09-543-681A-7696	Sequence 7696, Ap
11	89.5	6.6	281	4	US-09-198-452A-749	Sequence 749, App
12	89.5	6.6	543	4	US-09-328-352-5845	Sequence 5845, Ap
13	89.5	6.6	546	4	US-09-345-236B-98	Sequence 98, Appl
14	89.5	6.6	546	4	US-09-345-236B-121	Sequence 121, App
15	89.5	6.6	560	4	US-09-446-301A-50	Sequence 50, Appl
16	88.5	6.6	552	4	US-09-446-301A-4	Sequence 4, Appli
17	88.5	6.6	552	4	US-09-099-932-4	Sequence 4, Appli
18	88	6.5	488	4	US-09-489-039A-13363	Sequence 13363, A
19	87.5	6.5	732	2	US-08-533-669A-18	Sequence 18, Appl
20	87.5	6.5	732	4	US-09-307-143-4	Sequence 4, Appli
21	87.5	6.5	732	4	US-09-183-861-18	Sequence 18, Appl
22	87.5	6.5	732	4	US-09-022-765-18	Sequence 18, Appl
23	87.5	6.5	732	4	US-09-551-974A-18	Sequence 18, Appl
24	87.5	6.5	732	4	US-09-565-501A-18	Sequence 18, Appl
25	87.5	6.5	732	4	US-09-639-206A-18	Sequence 18, Appl
26	87.5	6.5	732	4	US-09-874-923-18	Sequence 18, Appl
27	87	6.4	491	3	US-09-029-267-2	Sequence 2, Appli
28	86.5	6.4	331	4	US-09-134-001C-3626	Sequence 3626, Ap
29	84.5	6.3	528	4	US-08-637-670-27	Sequence 27, Appl
30	84.5	6.3	691	4	US-09-830-217-12	Sequence 12, Appl
31	84	6.2	492	4	US-09-107-532A-7062	Sequence 7062, Ap
32	83.5	6.2	264	2	US-08-719-697-8	Sequence 8, Appli
33	83.5	6.2	264	3	US-08-727-616A-8	Sequence 8, Appli
34	83.5	6.2	264	4	US-09-481-756-8	Sequence 8, Appli
35	83.5	6.2	286	1	US-07-721-775A-2	Sequence 2, Appli
36	83.5	6.2	286	1	US-08-339-658-2	Sequence 2, Appli
37	83.5	6.2	286	3	US-09-263-933-7	Sequence 7, Appli
38	83.5	6.2	286	3	US-09-263-933-14	Sequence 14, Appl
39	83.5	6.2	286	3	US-09-263-933-21	Sequence 21, Appl
40	83.5	6.2	286	4	US-09-025-769B-265	Sequence 265, App
41	83.5	6.2	286	4	US-09-025-769B-362	Sequence 362, App
42	83.5	6.2	286	4	US-09-919-901-7	Sequence 7, Appli
43	83.5	6.2	286	4	US-09-919-901-14	Sequence 14, Appl
44	83.5	6.2	286	4	US-09-919-901-21	Sequence 21, Appl
45	83.5	6.2	299	4	US-09-025-769B-285	Sequence 285, App

ALIGNMENTS

RESULT 1
US-09-252-991A-29841
; Sequence 29841, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29841
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29841

Alignment Scores:
Pred. No.: 4.77e-34 Length: 273
Score: 364.00 Matches: 85
Percent Similarity: 55.83% Conservative: 49
Best Local Similarity: 35.42% Mismatches: 88
Query Match: 26.98% Indels: 18
DB: 4 Gaps: 4

US-09-732-091-3 (1-759) x US-09-252-991A-29841 (1-273)

Qy 28 GAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTTTTT 87
Db 28 GlnLeuLeuGluArgValSerAsnAspAspLeuGluProLeuValGluTyrIleLeu--- 46

```
Qy         88 GGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATACAAAAGG 147
                                |||       |||  |||:::     :::::::::||||||
Db         47 ---------------LysAlaArgThrGluSerLeuSerLysGlnValAspPheLysArg 61

Qy        148 CATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGG 207
                  ::::::   :::||||||       |||  :::||||||:::   :::|||
Db         62 TrpHisProGluHisArgArgTyrAlaSerAlaIleLeuAspGluLeuArgLeuPheGly 81

Qy        208 AGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGC 267
                ||||||||||||:::     :::      |||      |||   |||::::::
Db         82 GlyAsnSerPheAlaAsnLeuTrpArgLysSerGlyProSerTyrThrGluValValArg 101

Qy        268 GATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATTGAA 327
                ||||||     ||||||||||||||                ||||        :::|||
Db        102 AspValAlaGlyLysLeuLysValLysGlyValGlySerMetGluLeuIleGluLeuGlu 121

Qy        328 CAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTG 387
                :::   |||:::      ||||||   ::::::||||||:::          |||:::
Db        122 GluAlaMetValGlnSerIleLeuArgGlnAlaLeuGluLysSerSerGlyGluAspArg 141

Qy        388 AAAGAAATGTGCGATGAATTA------------------TCCATAAAAAACACGGACAAT 429
                :::|||:::   :::   |||                          |||   |||
Db        142 ArgGluLeuGluGluIleLeuArgGluAlaGlyLeuAspLysThrLysMetThrAlaLeu 161

Qy        430 TTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAA 489
                ||||||      ||||||||          |||  :::    :::
Db        162 LeuAsnGlySerAlaLeuSerGly------LeuValValProAlaValAlaArgMetIle 179

Qy        490 TCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGT 549
                  |||:::   :::   ::::::   |||::::::|||:::   :::||||||   |||
Db        180 LeuTyrArgThrSerThrValIleValAsnSerMetAlaGlnGlnLeuLeuGlyHisGly 199

Qy        550 TTATCGCTTGCG------------GGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTA 597
                |||      |||            |||                |||   ::::::   |||
Db        200 LeuArgSerAlaValValAlaGlyGlyThrPheAlaGlyGlyArgAlaValAlaAlaLeu 219

Qy        598 ACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCG 657
                ||||||||||||||||:::|||   ||||||||||||||:::|||:::|||||||||
Db        220 AlaGlyProValGlyTrpValIleAlaGlyValTrpThrAlaValAspLeuAlaGlyPro 239

Qy        658 GCTTATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAG 717
                ||||||||||||||||||||      :::   :::|||   ||||||||||   :::
Db        240 AlaTyrArgValThrIleProCysValLeuHisIleAlaMetLeuArgLeuLysAlaArg 259
```

RESULT 2
US-09-252-991A-29840
; Sequence 29840, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29840
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29840

Alignment Scores:

Pred. No.:	5.79e-32	Length:	321
Score:	346.50	Matches:	82
Percent Similarity:	51.88%	Conservative:	42
Best Local Similarity:	34.31%	Mismatches:	112
Query Match:	25.69%	Indels:	3
DB:	4	Gaps:	3

US-09-732-091-3 (1-759) x US-09-252-991A-29840 (1-321)

```
Qy          1 ATGGCATACAAATATGATAGAGACTTG---GAATTTTTAAAGCAATTGGAATCTAGTGAT 57
              ||||||      :::|||   ||||||   :::   |||             |||   |||
Db         53 MetAlaIleIleHisAspAlaAspLeuAlaAspValLeuLeuSerAlaSerSerAspAsp 72

Qy         58 TTATTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAA 117
              :::      |||  :::|||:::             :::|||           :::
Db         73 IleArgLeuLeuIleAspValIleThrAspAsnGlyAsnGlyArgIleSerLeuSerSer 92

Qy        118 AAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAA 177
              :::      :::      :::   |||
Db         93 SerValCysArgGlnLeuSerAlaAlaLysGluGlyValValGlyGluPheGluArgGly 112

Qy        178 AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAA--- 234
              :::|||||||||||||      :::|||   ||||||      :::      :::
Db        113 MetValAlaGluGluLeuMetArgPheGlyGlyAsnSerLeuMetAsnLeuPheArgGly 132

Qy        235 GGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAAT 294
              |||   ||||||   ||||||||||:::|||   ||||||          :::   |||:::
Db        133 GlySerGlyValProTyrLysGluLeuLeuSerAspValAlaSerHisValGlyValSer 152

Qy        295 TACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAA 354
                          :::      :::|||      ::::::::::|||:::   |||
Db        153 LysGluSerThrGlyAspCysAlaArgMetGluMetAlaIleIleThrLysValPheGlu 172

Qy        355 AGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATA 414
              :::|||:::      |||   :::|||:::                :::   |||
Db        173 GlnSerIleGlyArgMetSerGluGluAspLysAlaThrPhePheGluSerLeuGlyThr 192

Qy        415 AAAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAA 474
                          :::      ||||||   |||||   :::
Db        193 SerTyrArgSerGlyMetGlyProValAlaLeu---AlaAlaLeuIleAlaSerLeuSer 211

Qy        475 ATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACC 534
                 |||      ||||||   ||||||   :::||||||:::|||          :::
Db        212 AlaSerGlyMetThrSerTyrGlyLeuAlaAlaMetValAlaSerAlaThrMetSerSer 231

Qy        535 ATTCTAGGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTT 594
              ::::::||||||||:::::::||||||||          |||   |||   ||||||
Db        232 LeuValGlyArgGlyValAlaLeuAlaGlyGlyAlaThrLeuGlyArgGlyLeuSerVal 251

Qy        595 TTAACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGG 654
              ||||||||||||||||||||      |||   |||:::|||||||||   |||:::|||
Db        252 LeuThrGlyProValGlyTrpAlaIleAlaGlyIleTrpThrAlaPheAspLeuAlaSer 271

Qy        655 CCGGCTTATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAA 711
              ||||||||||||||||||:::|||      |||   :::      :::|||   |||
Db        272 ProAlaTyrArgValThrLeuProCysValIleGlnIleGlyHisMetArgGlnLys 290
```

RESULT 3
US-09-489-039A-8852
; Sequence 8852, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8852
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8852

Alignment Scores:

Pred. No.:	2.52e-20	Length:	258

Score:	247.00	Matches:	70
Percent Similarity:	47.58%	Conservative:	48
Best Local Similarity:	28.23%	Mismatches:	100
Query Match:	18.31%	Indels:	30
DB:	4	Gaps:	6

US-09-732-091-3 (1-759) x US-09-489-039A-8852 (1-258)

```
Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
Db         14 TyrLeuGluAspAspAspLeuAlaPheLeuProGluCysSerGluAlaHisLeuGluAla 33

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
Db         34 PheThrArgIleLeuThrHisGlyGluAsnGlyLysProArgLeuSerSerThrLeuLeu 53

Qy        127 AGCTCC---------ATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAA 177
Db         54 ArgAsnGluThrPheLeuAlaMetGluGlyHisProGluArgTyrArgArgAsnTrpGln 73

Qy        178 AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGC 237
Db         74 LeuIleAlaGlyGluLeuGlnHisPheGlyGlyAspSerIleAlaAsnThrLeuArgArg 93

Qy        238 GAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTAC 297
Db         94 HisGlyLysPheTyrArgAlaIleLeuLeuAspValCysLysArgLeuLysAlaLysVal 113

Qy        298 AACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGA 357
Db        114 AspLysGlnLeuSerThrProGlnIleGluGlnGlnLeuLeuAlaHisPheLeuGlnHis 133

Qy        358 AGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAA 417
Db        134 SerTrpAsnLysLeuAsnAlaGluGlnLysAlaGlnPheLeuAlaAlaValGluCysArg 153

Qy        418 AAC------ACGGACAATTTA----------------AACAGACAAGCCTTAAGCGCG 453
Db        154 SerHisGluLeuAspSerLeuMetAlaHisLeuLeuArgHisArgLysLeuSerGluGly 173

Qy        454 GCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTT 513
Db        174 ValThrLeuLeuLeu--------------------AspGluArgLeuThrAlaIleLeu 186

Qy        514 GCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGCAATCAGGTG 573
Db        187 ArgThrHisAlaAlaValSerValIleGlyHisGlyLeuValArgGlyAlaGly------ 204

Qy        574 CTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGG 633
Db        205 --------------------LeuGlyGlyProLeuGlyAlaAlaLeuAsnSerValLys 217

Qy        634 ACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGCATTGTGGTT 693
Db        218 Ala---------ValSerGlySerAlaTyrArgValThrIleProAlaValLeuHisIle 234

Qy        694 GCCACTTTACGCCTAAAAACACAG 717
Db        235 AlaCysLeuArgGlnMetLeuGln 242
```

RESULT 4
US-09-543-681A-6437
; Sequence 6437, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6437
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6437

Alignment Scores:

Pred. No.:	0.00316	Length:	739
Score:	103.00	Matches:	57
Percent Similarity:	45.81%	Conservative:	47
Best Local Similarity:	25.11%	Mismatches:	72
Query Match:	7.64%	Indels:	51
DB:	4	Gaps:	10

US-09-732-091-3 (1-759) x US-09-543-681A-6437 (1-739)

```
Qy         16 GATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAG 75
Db        517 AspLysValSerAspPheLeuValSerLeuValAsnSerSerLeuAspGlyLeuLysGlu 536

Qy         76 GTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATA 135
Db        537 Leu---------GlyLeuAsp------LysLysIleIleGluGluMetLysAsnSerIle 551

Qy        136 GAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTG 195
Db        552 Gln-----------------AspLysLeuLysMetAlaThrLysIleLeuValThrVal 565

Qy        196 CAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAA 255
Db        566 ValLeuPheValAlaAlaThrAlaLeuSerPheValIleGlyProAlaMet--------- 582

Qy        256 GAGATTTTATGCGATGTGTGCGATAAA--------------------------TTAAAG 288
Db        583 LysGlyIleSerAspAlaValAsnLysIleSerAsnGlnGlnIleArgGlnIleLeuLys 602

Qy        289 GTCAATTACAACAAGAAAACTGAAACGACTTTA------ATTGAACAAAACATGCTTTCT 342
Db        603 LysValLeuAsnAspGlyLeuGluAlaValLeuGlyLysMetIleLysAspIleIleIle 622

Qy        343 AAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGAT 402
Db        623 LysAlaLeuGluGluAlaLeuGluLysIleAsp-----------LysGlnLeuAlaLys 638

Qy        403 GAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCCTTA--------------- 447
Db        639 GluIleSerLysLysAlaSerIleMetLeuAsnArgThrValValAlaSerLysLeuThr 658

Qy        448 ---AGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCT 504
Db        659 AsnSerAlaAlaThrAsnThrValAsnIleTyrGlySer--------------------- 671

Qy        505 GTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGC 564
Db        672 ---ValIleAlaSerLysIleIleGln--------------------SerIleAlaGly 683

Qy        565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACA 624
Db        684 SerLysLysLeuThrAlaValLeuAspIleIle-GlnLysLeuMetAspLysIleMetGl 703

Qy        625 GGCGTATGGACAGCGATTG 643
Db        703 uThrTyrHisGluAsnIle 709
```

RESULT 5
US-08-658-639-12
; Sequence 12, Application US/08658639
; Patent No. 5914238
; GENERAL INFORMATION:
; APPLICANT: KEESEE, SUSAN
; APPLICANT: OBAR, ROBERT
; APPLICANT: WU, YING-JYE

```
;    TITLE OF INVENTION:  MATERIALS AND METHODS FOR DETECTION OF
;    TITLE OF INVENTION:  BREAST CANCER
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Testa, Hurwitz & Thibeault
;      STREET:  125 High St.
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,639
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, THOMAS C
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  MTP-021 (8395/24)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  173 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-658-639-12

Alignment Scores:
Pred. No.:                0.00254        Length:          173
Score:                    101.50         Matches:         41
Percent Similarity:       43.58%         Conservative:    37
Best Local Similarity:    22.91%         Mismatches:      72
Query Match:              7.52%          Indels:          29
DB:                       2              Gaps:            5

US-09-732-091-3 (1-759) x US-08-658-639-12 (1-173)

Qy         26 TGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTTT 85
              |||          :::   |||   |||          |||
Db          2 TrpProSerLysAlaArgTrpMetLeuPheAlaSerTrp---------------------- 14

Qy         86 TTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATACAAAA 145
                                              |||          ||||||          :::
Db         15 --------------------------GlnLysThrTrpValAlaProGlyTyrValArg 25

Qy        146 GGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATG 205
                  ::::::      |||   |||      |||||||||  :::|||
Db         26 LysPheValLeuMetArgAlaAsnIleGlnAlaValSerLeuLys--IleGlnThrLeuL 45

Qy        206 GGAGC---AATAGTTTTGCGAGTTTCATTAAAGGC-----------GAAGGAGTCTTAT 250
                |||   ||||||   |||      :::||||||               |||   :::
Db         45 ysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAlaMetGlyThrMetA 65

Qy        251 ACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTG 310
                :::::::::          |||:::  :::  :::  ::::::   |
Db         65 snArgGlnLeuLysLeuProGlnIleGlnLysIleMetMetGluPheGluArgGlnAlaG 85

Qy        311 AAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAA 370
                ||         :::::::::   |||::::::      ::::::   ::::::   :::
Db         85 luIleMetAspMetLysGluGluMetMetAsnAspAlaIleAspAspAlaMetGlyAspG 105

Qy        371 TGGATGATGAAGAA--------------GTGAAAGAAATGTGCGATGAATTATCCATAA 415
                :::|||||||||              |||   ::::::   |||||||||   :::
Db        105 luGluAspGluGluGluSerAspAlaValValSerGlnValLeuAspGluLeuGlyLeuS 125

Qy        416 AAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAA 475
                |||||   |||:::          :::      :::|||::::::
Db        125 erLeuThrAspGluLeuSerAsnLeuProSerThrGlyGlySerLeuServal---AlaA 144

Qy        476 TGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCA 528
                ||||||   |||:::          |||      :::|||:::|||   |||
Db        144 laGlyGlyLysLysAlaGluAlaAlaAlaSerAlaLeuAlaAspAlaAspAla 161

RESULT 6
US-08-944-604-12
; Sequence 12, Application US/08944604
; Patent No. 6218131
;  GENERAL INFORMATION:
;    APPLICANT:  KEESEE, SUSAN
;    APPLICANT:  OBAR, ROBERT
;    APPLICANT:  WU, YING-JYE
;    TITLE OF INVENTION:  MATERIALS AND METHODS FOR DETECTION OF
;    TITLE OF INVENTION:  BREAST CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Testa, Hurwitz & Thibeault
;      STREET:  125 High St.
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/944,604
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, THOMAS C
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  MTP-021 (8395/24)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  173 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-944-604-12

Alignment Scores:
Pred. No.:                0.00254        Length:          173
Score:                    101.50         Matches:         41
Percent Similarity:       43.58%         Conservative:    37
Best Local Similarity:    22.91%         Mismatches:      72
Query Match:              7.52%          Indels:          29
DB:                       3              Gaps:            5

US-09-732-091-3 (1-759) x US-08-944-604-12 (1-173)

Qy         26 TGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTTT 85
              |||          :::   |||   |||          |||
Db          2 TrpProSerLysAlaArgTrpMetLeuPheAlaSerTrp---------------------- 14

Qy         86 TTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATACAAAA 145
                                              |||          ||||||          :::
Db         15 --------------------------GlnLysThrTrpValAlaProGlyTyrValArg 25

Qy        146 GGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATG 205
                  ::::::      |||   |||      |||||||||  :::|||
Db         26 LysPheValLeuMetArgAlaAsnIleGlnAlaValSerLeuLys--IleGlnThrLeuL 45

Qy        206 GGAGC---AATAGTTTTGCGAGTTTCATTAAAGGC-----------GAAGGAGTCTTAT 250
                |||   ||||||   |||      :::||||||               |||   :::
```

```
Db          45 ysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAlaMetGlyThrMetA 65

Qy         251 ACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTG 310
                 :::::::::                  |||:::   :::   :::   ::::::   |
Db          65 snArgGlnLeuLysLeuProGlnIleGlnLysIleMetMetGluPheGluArgGlnAlaG 85

Qy         311 AAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAA 370
                ||         :::::::::   |||::::::      ::::::   ::::::   :::
Db          85 luIleMetAspMetLysGluGluMetMetAsnAspAlaIleAspAspAlaMetGlyAspG 105

Qy         371 TGGATGATGAAGAA--------------GTGAAAGAAATGTGCGATGAATTATCCATAA 415
                 :::|||||||||               |||  ::::::   |||||||||   :::
Db         105 luGluAspGluGluGluSerAspAlaValValSerGlnValLeuAspGluLeuGlyLeuS 125

Qy         416 AAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAA 475
                   ||||||   |||:::            :::      :::|||::::::
Db         125 erLeuThrAspGluLeuSerAsnLeuProSerThrGlyGlySerLeuSerVal---AlaA 144

Qy         476 TGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCA 528
                 ||||||   |||:::          |||      :::|||:::|||   |||
Db         144 laGlyGlyLysLysAlaGluAlaAlaAlaSerAlaLeuAlaAspAlaAspAla 161

RESULT 7
US-08-944-604-16
; Sequence 16, Application US/08944604
; Patent No. 6218131
;  GENERAL INFORMATION:
;    APPLICANT:  KEESEE, SUSAN
;    APPLICANT:  OBAR, ROBERT
;    APPLICANT:  WU, YING-JYE
;    TITLE OF INVENTION:  MATERIALS AND METHODS FOR DETECTION OF
;    TITLE OF INVENTION:  BREAST CANCER
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Testa, Hurwitz & Thibeault
;      STREET:  125 High St.
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/944,604
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MEYERS, THOMAS C
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  MTP-021 (8395/24)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  222 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-944-604-16

Alignment Scores:
Pred. No.:                0.00555        Length:          222
Score:                    99.00          Matches:         39
Percent Similarity:       47.80%         Conservative:    48
Best Local Similarity:    21.43%         Mismatches:      83
Query Match:              7.34%          Indels:          12
DB:                       3              Gaps:            5

US-09-732-091-3 (1-759) x US-08-944-604-16 (1-222)
```

```
Qy          16 GATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAG 75
                |||||:::   :::   |||:::      |||          :::   |||:::   :::
Db          30 AspArgGluArgGlnLysLeuGluThrGlnGluLysLysIleIleAlaAspIleLysLys 49

Qy          76 GTGCTTGTTTTTGGTAAA---GACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCC 132
                :::          |||:::   |||      :::          :::   |||      :::
Db          50 MetAlaLysGlnGlyGlnMetAspAlaValArgIleMetAlaLysAspLeuValArgThr 69

Qy         133 ATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAG 192
                    |||   |||                    :::   ::::::   :::
Db          70 ArgArgTyrValArgLysPheValLeuMetArgAlaAsnIleGlnAlaValSerLeuLys 89

Qy         193 TTGCAATACTATGGGAGC---AATAGTTTTGCGAGTTTCATTAAAGGC------------ 237
                :::|||          |||   ||||||   |||       :::||||||
Db          90 IleGlnThrLeuLysSerAsnAsnSerMetAlaGlnAlaMetLysGlyValThrLysAla 109

Qy         238 GAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTAC 297
                |||   :::   :::::::::                        |||:::   :::   :::
Db         110 MetGlyThrMetAsnArgGlnLeuLysLeuProGlnIleGlnLysIleMetMetGluPhe 129

Qy         298 AACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGA 357
                ::::::   |||          :::::::::   |||::::::      ::::::
Db         130 GluArgGlnAlaGluIleMetAspMetLysGluGluMetMetAsnAspAlaIleAspAsp 149

Qy         358 AGTTTGGAAGAAATGGATGATGAAGAA--------------GTGAAAGAAATGTGCGAT 402
                ::::::   :::   :::|||||||||               |||   ::::::   |||
Db         150 AlaMetGlyAspGluGluAspGluGluGluSerAspAlaValValSerGlnValLeuAsp 169

Qy         403 GAATTATCCATAAAAAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTA 462
                ||||||   :::       ||||||   |||:::             :::      :::|||
Db         170 GluLeuGlyLeuSerLeuThrAspGluLeuSerAsnLeuProSerThrGlyGlySerLeu 189

Qy         463 ACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCG 522
                ::::::          ||||||   |||:::          |||      :::|||:::|||
Db         190 SerVal---AlaAlaGlyGlyLysLysAlaGluAlaAlaAlaSerAlaLeuAlaAspAla 208

Qy         523 GTCGCA 528
                |||
Db         209 AspAla 210

RESULT 8
US-09-134-001C-5279
; Sequence 5279, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCU
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5279
;   LENGTH: 274
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5279

Alignment Scores:
Pred. No.:                0.00608        Length:          274
Score:                    99.00          Matches:         62
Percent Similarity:       40.77%         Conservative:    44
Best Local Similarity:    23.85%         Mismatches:      88
Query Match:              7.34%          Indels:          66
DB:                       4              Gaps:            14

US-09-732-091-3 (1-759) x US-09-134-001C-5279 (1-274)
```

```
Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
              |||   |||:::   ||||||
Db          4 TyrAsnTyrAsnAsnAspLeu---------------------------------- 10

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
                    |||:::      |||  |||::::::  :::  :::  :::
Db         11 ---------PheLeuIleLeuLysLysPheGlyAspGluMetAsnLysAspThrIleGlu 27

Qy        127 AGCTCC-----------------------------------------------ATAGAA 138
              :::|||                                               :::|||
Db         28 ThrSerArgGluAlaLeuIleLysAspValValMetIleAlaAlaArgIleLeuLeuGlu 47

Qy        139 TACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAA 198
                  |||                  |||||||||  :::|||
Db         48 SerGlyAlaGluGlyThrArgValGluAspThrMetAlaArgIleAlaThrLysLeuGly 67

Qy        199 TACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAG 258
              |||      ||||||||||||  :::          |||  ||||||:::  |||
Db         68 TyrProGluSerAsnSerPheValThrAsnThrValIleGluPheValLeuHisAsnGlu 87

Qy        259 ATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACT---GAAACG 315
                  :::                  |||      |||||  :::|||
Db         88 AlaTyrProArgLeu---------------------TyrArgIleLysThrArgAspThr 100

Qy        316 ACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAA---GAAATG 372
                |||||:::  :::      ::::::|||      :::              |||
Db        101 AsnLeuIleLysIleSerGlnAlaAsnGluIleSerArgGlnIleThrAsnGlyThrMet 120

Qy        373 GATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATA---AAAAACACGGACAAT 429
                |||||  |||          :::|||:::  :::  |||          :::
Db        121 ThrLeuGluGluAlaLysTyrGlnLeuGluGluIleTyrValAlaLysArgAspSerSer 140

Qy        430 TTAAACAGACAAGCCTTAAGCGCG---------GCGACTTTAACGCTGTTTAAAATGGGG 480
              |||      :::   ::::::|||        ||||||      |||:::      |||
Db        141 LeuProPheLysGlyIleAlaAlaAlaIleIleAlaThrSerPheLeuTyrLeuGlnGly 160

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              |||              :::|||      |||:::          :::|||  ||||||
Db        161 Gly------------ArgLeuValAspIleIleThrAlaValLeuAlaGlyThrIle--- 175

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTG------AGCTTT 594
                  |||      :::          ::::::|||   |||  |||      |||
Db        176 ------GlyTyrLeuValVal------GluIleLeuAspArgLysLeuHisAlaGlnPhe 191

Qy        595 TTAACAGGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGG 654
              :::        :::|||   ::::::   |||:::          |||   :::   |||
Db        192 IleProGluPheIleGlySerLeuValIleGlyIle---------IleSerValIleGly 208

Qy        655 CCGGCTTATAGGGTAACCATACCG------------GCATGCATTGTGGTTGCCACTTTA 702
                |||:::        :::|||            |||   |||::::::|||   :::
Db        209 HisAlaPhe---------ValProSerGlyAspLeuAlaThrIleIleIleAlaAlaVal 225
```

RESULT 9
US-09-919-172-29
; Sequence 29, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 2125
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673545 3774181CD1
US-09-919-172-29

Alignment Scores:

Pred. No.:	0.0431	Length:	2125
Score:	95.00	Matches:	46
Percent Similarity:	44.75%	Conservative:	35
Best Local Similarity:	25.41%	Mismatches:	62
Query Match:	7.04%	Indels:	38
DB:	4	Gaps:	9

US-09-732-091-3 (1-759) x US-09-919-172-29 (1-2125)

```
Qy          4 GCATACAAATATGATAGAGACTTGGAATTTTTAAAG------CAATTGGAATCTAGTGAT 57
              :::::::::  :::::::::||||||  |||:::      |||  :::|||
Db        986 SerPheArgAspGluLysGluLeuGluArgLeuGlnIleCysGlnArgLysSerAspHis 1005

Qy         58 TTATTGGATTTGTTT------------GAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAA 105
              |||   :::  |||            |||  |||:::           ||||||
Db       1006 LeuLysGluGlnPheGluLysSerHisGluGlnLeuLeuGlnAsnIleLysAlaGluLys 1025

Qy        106 AGACACAATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCT 165
                :::|||:::|||:::          |||                 ::::::  |||
Db       1026 GluAsnAsnAspLysIleGlnArgLeuAsnGluGluLeuGluLysSerAsnGluCysAla 1045

Qy        166 AAATACGCAGAAAGAATCGCTGAAGAGTTG------CAATACTATGGGAGCAATAGTTTT 219
                :::      ::::::      |||||||||      |||         :::
Db       1046 GluMetLeuLysGlnLysValGluGluLeuThrArgGlnAsnAsnGluThrLysLeuMet 1065

Qy        220 GCGAGTTTCATTAAAGGCGAAGGA---------GTCTTATACAAAGAGATTTTATGCGAT 270
                  |||:::   |||            ||||||   |||:::   :::
Db       1066 MetGlnArgIleGlnAlaGluSerGluAsnIleValLeuGluLysGlnThrIleGlnGln 1085

Qy        271 GTGTGCGATAAATTAAAGGTC------------------------------------ 291
                |||:::   ||||||:::
Db       1086 ArgCysGluAlaLeuLysIleGlnAlaAspGlyPheLysAspGlnLeuArgSerThrAsn 1105

Qy        292 ---AATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATC 348
                :::   :::|||:::|||:::|||          ||||||:::          ||||||
Db       1106 GluHisLeuHisLysGlnThrLysThr---------GluGlnAspPheGlnArgLysIle 1122

Qy        349 ------TTAGAAAGAAGTTTGGAAGAAATGGATGAT------GAAGAAGTGAAAGAAATG 396
                   ||||||      |||   :::      :::          |||   |||:::
Db       1123 LysCysLeuGluGluAspLeuAlaLysSerGlnAsnLeuValSerGluPheLysGlnLys 1142

Qy        397 TGCGATGAATTATCCATA------------------AAAAACACGGACAATTTAAACAGA 438
                |||||:::  :::|||                 |||          |||||||||
Db       1143 CysAspGlnGlnAsnIleIleIleGlnAsnThrLysLysGluValArgAsnLeuAsnAla 1162

Qy        439 CAA 441
                :::
Db       1163 Glu 1163
```

RESULT 10
US-09-543-681A-7696
; Sequence 7696, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7696
; LENGTH: 1086
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7696

```
Alignment Scores:
Pred. No.:                0.0833          Length:          1086
Score:                    91.50           Matches:         43
Percent Similarity:       36.65%          Conservative:    38
Best Local Similarity:    19.46%          Mismatches:      73
Query Match:              6.78%           Indels:          67
DB:                       4               Gaps:            9

US-09-732-091-3 (1-759) x US-09-543-681A-7696 (1-1086)

Qy          7 TACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGAT 66
              :::   |||||||||       |||   ::::::|||::::::
Db        786 PheAlaTyrAspArgGlyLysGlu---IleGluGlnMetGln------------------ 798

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACC 126
                ||||||      :::   ||||||          :::         |||||||||
Db        799 ---PheGluThrThrLeuIleGlyLys------SerArgAlaGluGlnGluLysLeuAsn 815

Qy        127 AGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCT 186
              :::
Db        816 Ala---------------------------------------------------------- 816

Qy        187 GAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTC 246
                                                      ::::::         |||
Db        817 ----------------------------------------LeuArgGlnIleAspVal 822

Qy        247 TTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAA 306
              ||||||::::::        |||:::   :::|||      ||||||      :::
Db        823 LeuTyrGlnGlnAlaSerValAspLeuGlyGluLysGluLeuValAsnLeuGlnArgAsn 842

Qy        307 ACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAA 366
              |||   |||       :::|||   :::      ::::::   :::|||
Db        843 ValGluLeuThr------LysGlnGlnIleGluGluGluLeuArgLysArgGluAlaMet 860

Qy        367 GAAATGGAT----------------------------------GATGAAGAAGTGAAA 390
              :::   |||                                  |||
Db        861 LysGlyAspProMetAlaGlyLeuLysGlnGlyLeuSerAspPheSerGluSerAlaMet 880

Qy        391 GAAATGTGCGATGAATTATCCATAAAAAACACGGACAATTTAAACAGACAAGCCTTAAGC 450
              ::::::   :::   :::          |||:::   ||||||      :::   :::
Db        881 AspValMetGluAsnValArgAsnValThrThrAsnAlaLeuAsnAsnMetSerAspAla 900

Qy        451 GCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATT 510
              |||          |||          |||   ||||||   :::         |||||||||
Db        901 LeuAlaAspPheAlaLeuThrGlyLysGlySerPheLysAspPheAlaAsnAlaValIle 920

Qy        511 ------GTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCGCTTGCGGGC 564
                    :::       |||  |||   ::::::  :::   :::         |||
Db        921 SerAspIleThrArgMetValMetLysMetLeuIlePheLysAlaIleGluAlaGlyGly 940

Qy        565 AATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATTACA 624
                          :::   :::   |||           :::||||||:::
Db        941 --------------GlnAlaMetGlyPhe---------AspMetGlyTrpMetSerLys 952

Qy        625 GGC 627
              |||
Db        953 Gly 953

RESULT 11
US-09-198-452A-749
; Sequence 749, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 749
;   LENGTH: 281
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
US-09-198-452A-749

Alignment Scores:
Pred. No.:                0.0801          Length:          281
Score:                    89.50           Matches:         56
Percent Similarity:       41.18%          Conservative:    49
Best Local Similarity:    21.96%          Mismatches:      103
Query Match:              6.63%           Indels:          47
DB:                       4               Gaps:            11

US-09-732-091-3 (1-759) x US-09-198-452A-749 (1-281)

Qy         64 GATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTG 123
              |||       :::|:|   ||||||      :::|||::::::        ||||||
Db         30 AspSerArgAspValLysValPheSerLeuGluGlyLysGlnThrArgGlnGluLysThr 49

Qy        124 ACCAGCTCC---------------ATAGAATACAAAAGGCATGGCGATGATTACGCTAAA 168
              |||||||||                  |||   ::::::      |||:::      :::
Db         50 ThrSerSerLysGlyAsnThrArgThrGluSerArgLysPheAlaAspGluGluLysArg 69

Qy        169 TACGCAGAAAGAATCGCTGAA-----------------------GAGTTGCAATACTAT 204
              :::   |||||||||                           |||   |||   :::
Db         70 ValAspAspGluIleAlaGluValGlySerLysGluGluGluGlnGluSerGlnGluPhe 89

Qy        205 ------GGGAGCAATAGTTTTGCG------AGTTTCATTAAAGGCGAAGGAGTCTTATAC 252
                    |||:::||||||      |||   |||
Db         90 CysLeuAlaGluAsnAlaPheAlaGlyMetSerLeuIleAspIleAlaAlaAlaGlySer 109

Qy        253 AAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAA 312
                |||   :::   :::|||         :::   |||:::                :::
Db        110 AlaGluAlaValValGluValAla---ProIleAlaValSer---------SerIleAsp 125

Qy        313 ACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATG 372
              |||       ||||||      :::||||||   :::            :::   |||:::
Db        126 ThrGlnTrpIleGluAsnIleIleLeuSerThrValGluSerMetValIleSerGluIle 145

Qy        373 GATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACGGACAATTTA 432
              :::   |||:::::   |||:::          :::   :::      :::
Db        146 AsnGlyGluGlnLeuValGluLeu--------------ValLeuAspAlaSerSerSer 160

Qy        433 AACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCT 492
                  :::|||          |||   |||||||||   :::   |||
Db        161 ValProGluAlaPheValGlyAlaAsnLeuThrLeuValGlnSerGly------------ 176

Qy        493 TATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTA 552
                    |||:::|||       ::::::   |||       |||   :::
Db        177 GlnAspLeuSerValLysPheSerSerPheValAspAlaThrGlnMetAlaGluAlaAla 196

Qy        553 TCGCTTGCGGGCAAT------CAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGT--- 603
                |||       |||             |||:::      :::|||   |||   |||
Db        197 AspLeuValThrAsnAsnProSerGlnLeuSerSerLeuValSerAlaLeuLysGlyHis 216

Qy        604 --------------------CCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATT 642
                                   ||||||   ::::::      :::
Db        217 GlnLeuThrLeuLysGluPheSerValGlyAsnLeuLeuValGlnLeuProLysIleGlu 236

Qy        643 GATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTA 702
              ::::::      |||   :::                       ::::::   :::|||
Db        237 GluValGlnThrProLeuHis--------------------MetIleAlaSerThrIle 249

Qy        703 CGCCTAAAAACACAGCAAGCCAATGGAGATAAGAAGTCGTTGCAA 747
              |||   :::   ::::::         |||:::         |||
Db        250 ArgHisArgGluGluLysAspGlnArgAspGlnAsnGlnLysGln 264

RESULT 12
US-09-328-352-5845
; Sequence 5845, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
```

```
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5845
;   LENGTH: 543
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-5845

Alignment Scores:
Pred. No.:                0.106          Length:          543
Score:                    89.50          Matches:         40
Percent Similarity:       37.37%         Conservative:    31
Best Local Similarity:    21.05%         Mismatches:      52
Query Match:              6.63%          Indels:          67
DB:                       4              Gaps:            10

US-09-732-091-3 (1-759) x US-09-328-352-5845 (1-543)

Qy         16 GATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAG 75
Db        366 AspLysSerLeuGluGluLeuGluLysGluGluArgLysAspLeuLysAsnArgPheAsp 385

Qy         76 GTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATA 135
Db        385 ------------------------------------------------------------ 385

Qy        136 GAATACAAAAGGCATGGCGAT----------------------------GATTACGCT 165
Db        386 GluTyrLysAsnLysGlyGlnIleLysPheValThrPheHisGlnSerPheSerTyrGlu 405

Qy        166 AAATACGCAGAAAGAATC---GCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCG 222
Db        406 AspPheValGluGlyIleArgAlaGluThrValGluTyrSerAspGlyLysLysAsnIle 425

Qy        223 AGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGAT------------ 270
Db        426 GluTyrProValValSerGlyValPhe------LysLeuLeuCysAspThrAlaGlnSer 443

Qy        271 ------GTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATT 324
Db        444 LysValIleLeuGluSerGlnLysIleAsnPheAspSerAsnThr--------------- 458

Qy        325 GAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATG---GATGATGAA 381
Db        459 --------------AsnGluIleTrpLysMetSerLeuGlyArgAlaGlyGluAspGlu 473

Qy        382 GAAGTGAAAGAAATGTGC----------------------------------GATGAATTA 408
Db        474 AspIlePheAspTyrCysIleLysAsnHisCysValLeuLeuGlyTrpGlyAspGluLeu 493

Qy        409 TCCATAAAAAACACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTG 468
Db        494 AspPheSerAsnAla------ValAsnArgLysGlnIleGlu------------GluIle 507

Qy        469 TTTAAAATGGGGGGTTTTAAATCTTATCAA 498
Db        508 MetAspLysAsnGlyTyrGluAlaTyrArg 517

RESULT 13
US-09-345-236B-98
; Sequence 98, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
;  APPLICANT: Becnel, James J.
;  APPLICANT:  Tukuo, Fukuda
;  APPLICANT:  Moser, Bettina
;  APPLICANT:  Cockburn, Andrew
;  APPLICANT:  White, Susan E.
;  APPLICANT:  Undeen, Albert H.
;  TITLE OF INVENTION: No. 6521454el Baculoviruses,  Insecticidal
;  TITLE OF INVENTION:  Compositions, and Methods for Control of Invertebrates
;  FILE REFERENCE: 21042.0004
;  CURRENT APPLICATION NUMBER: US/09/345,236B
;  CURRENT FILING DATE:  1999-06-30
;  NUMBER OF SEQ ID NOS: 148
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 98
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: mosquito baculovirus
US-09-345-236B-98

Alignment Scores:
Pred. No.:                0.106          Length:          546
Score:                    89.50          Matches:         60
Percent Similarity:       36.86%         Conservative:    34
Best Local Similarity:    23.53%         Mismatches:      92
Query Match:              6.63%          Indels:          69
DB:                       4              Gaps:            10

US-09-732-091-3 (1-759) x US-09-345-236B-98 (1-546)

Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA------ 120
Db        153 ValPheGluValValGlyGlyGlyProGluGlyAspTyrAlaAlaGlyGluGluAspGlu 172

Qy        121 ---------------CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCT 165
Db        173 ValSerArgAsnSerLeuAsnPheAspMetAlaSerGluValGlnSerThrAspAlaAla 192

Qy        166 AAATACGCAGAA---------AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGT 216
Db        193 LysValMetGluLeuPheSerAlaLeuSerGluGluGlnArgAsnValIleLeuAsnAsn 212

Qy        217 TTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGC 276
Db        213 PheGlyAlaAlaProSerGlySerGlyThr---------------------------- 222

Qy        277 GATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATG 336
Db        223 ------------------------------ThrProProThrSerAlaGlnProAspMet 232

Qy        337 CTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATG 396
Db        233 -------------------------------GluValGluAspValGluThrValGluLys 242

Qy        397 TGCGATGAATTATCCATAAAAAACACGGACAATTTAAAC------------AGACAAGCC 444
Db        243 ProGluAsnLeuAsnAspIleIleThrAspGlnLeuArgAspPheMetAlaGlnGluLeu 262

Qy        445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG---------GGTTTTAAATCTTAT 495
Db        263 LysLysAlaAlaGluAsnTyrValProLysTrpGlySerThrValGlyGluSerLysSer 282

Qy        496 CAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCG 555
Db        283 AlaLeuAlaIleThrValAlaAspArgValSerArgSerPheMetTyrGluGlyArgIle 302

Qy        556 CTTGCGGGCAATCAGGTG--------------------------------CTTACAAGA 582
Db        303 ValAspTyrAsnGlnValValLeuHisIleLeuAspAsnTyrAspGlnArgLeuGluGlu 322

Qy        583 ACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATT---------------------- 621
Db        323 LeuLeuSerPheArgThrLysThr------TyrIleIleAlaGluGlyValProHisAsp 340

Qy        622 ACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCA 681
Db        341 SerLysValHisAspTyrValAspLeuThr------GlnTyrArgGluThrValProTyr 358

Qy        682 TGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAAGCCAAT 726
Db        359 SerIleAlaLeuAsnAsnLeuSerArgGlyValAspGlnAlaAsn 373
```

RESULT 14
US-09-345-236B-121
; Sequence 121, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
; APPLICANT: Becnel, James J.
; APPLICANT: Tukuc, Fukuda
; APPLICANT: Moser, Bettina
; APPLICANT: Cockburn, Andrew
; APPLICANT: White, Susan E.
; APPLICANT: Undeen, Albert H.
; TITLE OF INVENTION: No. 6521454el Baculoviruses, Insecticidal
; TITLE OF INVENTION: Compositions, and Methods for Control of Invertebrates
; FILE REFERENCE: 21042.0004
; CURRENT APPLICATION NUMBER: US/09/345,236B
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 121
; LENGTH: 546
; TYPE: PRT
; ORGANISM: mosquito baculovirus
US-09-345-236B-121

Alignment Scores:

Pred. No.:	0.106	Length:	546
Score:	89.50	Matches:	60
Percent Similarity:	36.86%	Conservative:	34
Best Local Similarity:	23.53%	Mismatches:	92
Query Match:	6.63%	Indels:	69
DB:	4	Gaps:	10

US-09-732-091-3 (1-759) x US-09-345-236B-121 (1-546)

```
Qy         67 TTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA------ 120
              :::|||||||||:::      |||   :::|||:::            |||:::
Db        153 ValPheGluValValGlyGlyGlyProGluGlyAspTyrAlaAlaGlyGluGluAspGlu 172

Qy        121 ---------------CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCT 165
                           |||         :::     :::             |||   |||
Db        173 ValSerArgAsnSerLeuAsnPheAspMetAlaSerGluValGlnSerThrAspAlaAla 192

Qy        166 AAATACGCAGAA---------AGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGT 216
              |||     |||           ::::::||||||   :::             |||:::
Db        193 LysValMetGluLeuPheAsnAlaLeuSerGluGluGlnArgAsnValIleLeuAsnAsn 212

Qy        217 TTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGC 276
              |||   :::         |||   |||
Db        213 PheGlyAlaAlaProSerGlySerGlyThr---------------------------- 222

Qy        277 GATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATG 336
                                       |||     |||      :::   :::|||
Db        223 ----------------------------ThrProProThrSerAlaGlnProAspMet 232

Qy        337 CTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATG 396
                                      |||::::::||   |||       |||
Db        233 -----------------------------GluValGluAspValGluThrValGluLys 242

Qy        397 TGCGATGAATTATCCATAAAAAACACGGACAATTTAAAC-----------AGACAAGCC 444
                 :::   |||:::         ||||||   |||             ::::::
Db        243 ProGluAsnLeuAsnAspIleIleThrAspGlnLeuArgAspPheMetAlaGlnGluLeu 262

Qy        445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG---------GGTTTTAAATCTTAT 495
              ||||||          :::   |||   |||         |||
Db        263 LysLysAlaAlaGluAsnTyrValProLysTrpGlySerThrValGlyGluSerLysSer 282

Qy        496 CAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCG 555
              ||||||:::   ||||||:::   |||:::::::::   :::      |||
Db        283 AlaLeuAlaIleThrValAlaAspArgValSerArgSerPheMetTyrGluGlyArgIle 302

Qy        556 CTTGCGGGCAATCAGGTG------------------------------CTTACAAGA 582
              :::      |||||||||                              |||
```

```
Db        303 ValAspTyrAsnGlnValValLeuHisIleLeuAspAsnTyrAspGlnArgLeuGluGlu 322

Qy        583 ACTCTGAGCTTTTTAACAGGTCCTGTTGGCTGGATCATT---------- 621
              |||||||||  |||          :::|||||
Db        323 LeuLeuSerPheArgThrLysThr------TyrIleIleAlaGluGlyValProHisAsp 340

Qy        622 ACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACCATACCGGCA 681
              :::   |||         :::|||:::           ||||||   |||:::|||
Db        341 SerLysValHisAspTyrValAspLeuThr------GlnTyrArgGluThrValProTyr 358

Qy        682 TGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAAGCCAAT 726
              |||   :::      |||                 |||||||||
Db        359 SerIleAlaLeuAsnAsnLeuSerArgGlyValAspGlnAlaAsn 373
```

RESULT 15
US-09-446-301A-50
; Sequence 50, Application US/09446301A
; Patent No. 6506893
; GENERAL INFORMATION:
; APPLICANT: EL SOLH, NEVINE
; APPLICANT: ALLIGNET, JEANINE
; TITLE OF INVENTION: POLYNUCLEOTIDES AND THEIR USE FOR DETECTING RESISTANCE
; TITLE OF INVENTION: TO STREPTOGRAMIN A OR TO STREPTOGRAMIN B AND RELATED
; TITLE OF INVENTION: COMPOUNDS
; FILE REFERENCE: 03715-0059
; CURRENT APPLICATION NUMBER: US/09/446,301A
; CURRENT FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Staphylococcus sp.
US-09-446-301A-50

Alignment Scores:

Pred. No.:	0.108	Length:	560
Score:	89.50	Matches:	37
Percent Similarity:	46.58%	Conservative:	31
Best Local Similarity:	25.34%	Mismatches:	47
Query Match:	6.63%	Indels:	31
DB:	4	Gaps:	7

US-09-732-091-3 (1-759) x US-09-446-301A-50 (1-560)

```
Qy         22 GACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTT 81
              |||::::::   ::::::  ||||||      :::||||||      :::|||  :::
Db        435 AspIleAspAlaValGluAlaLeuGlu------GluLeuLeuIleThrTyrGluGlyVal 452

Qy         82 GTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATAC 141
              |||          :::|||
Db        453 ValLeuPheAlaSerHisAspLys---------------------------------- 460

Qy        142 AAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATAC 201
                                    |||:::   :::   :::||||||:::|||
Db        461 ------------------------LysPheIleGlnAsnLeuAlaGluGlnLeuLeuIle 472

Qy        202 TATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATT 261
                 :::|||          |||         ||||||      |||   |||
Db        473 IleGluAsnAsnLysValLysLysPhe---------GluGlyThr---TyrIleGluTyr 488

Qy        262 TTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAA------ACTGAAACG 315
              |||       :::   ||||||  |||:::|||   |||:::|||          |||
Db        489 Leu---LysIleLysAspLysProLysLeuAsnThrAsnGluLysGluLeuLysGluLys 507

Qy        316 ACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATGGAT 375
                 :::|||   :::       :::|||   :::|||   :::          |||   :::
Db        508 LysMetIleLeuGluMetGlnIleSerSerLeuLeuSerLysIleSerMetGluGluAsn 527

Qy        376 GATGAAGAAGTGAAAGAAATGTGCGAT------GAATTATCCATAAAAAACACGGACAAT 429
                 :::|||:::   ||||||:::   :::       :::|||   :::|||          :::
Db        528 GluGluLysAsnLysGluLeuAspGluLysTyrLysLeuLysLeuLysGluLeuLysSer 547
```

```
Qy        430 TTAAACAGACAAGCCTTA 447
              ||||||:::      :::
Db        548 LeuAsnLysAsnIleMet 553
```

Search completed: July 5, 2004, 03:07:23
Job time : 21.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 5, 2004, 03:03:18 ; Search time 67.5 Seconds
(without alignments)
7000.427 Million cell updates/sec

Title: US-09-732-091-3
Perfect score: 1349
Sequence: 1 atggcatacaaatatgatag.....cgttgcaaatagaatccatt 759

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 2553080

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09732091/runat_02072004_182407_19065/app_query.fasta_1.903
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LCOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09732091_@CGN_1_1_35_@runat_02072004_182407_19065
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1279	94.8	253	9	US-09-732-091-4	Sequence 4, Appli
2	1279	94.8	265	9	US-09-732-091-44	Sequence 44, Appl
3	1270	94.1	253	12	US-10-335-977-9162	Sequence 9162, Ap
4	1270	94.1	253	12	US-10-335-977-9163	Sequence 9163, Ap
5	1270	94.1	256	12	US-10-335-977-9164	Sequence 9164, Ap
6	722	53.5	248	12	US-10-335-977-7699	Sequence 7699, Ap
7	716	53.1	237	12	US-10-335-977-7698	Sequence 7698, Ap
8	472.5	35.0	155	10	US-09-882-227-414	Sequence 414, App
9	252	18.7	49	9	US-09-732-091-20	Sequence 20, Appl
10	198	14.7	38	9	US-09-732-091-17	Sequence 17, Appl
11	194	14.4	41	9	US-09-732-091-19	Sequence 19, Appl
12	152	11.3	30	9	US-09-732-091-18	Sequence 18, Appl
13	150	11.1	30	9	US-09-732-091-16	Sequence 16, Appl
14	98.5	7.3	1009	12	US-10-282-122A-43832	Sequence 43832, A
15	97.5	7.2	815	16	US-10-437-963-191043	Sequence 191043,
16	97	7.2	916	12	US-10-282-122A-76490	Sequence 76490, A
17	96.5	7.2	373	15	US-10-369-493-42	Sequence 42, Appl
18	96.5	7.2	701	14	US-10-032-585-7400	Sequence 7400, Ap
19	96.5	7.2	1102	14	US-10-156-761-14395	Sequence 14395, A
20	95	7.0	517	12	US-10-282-122A-54543	Sequence 54543, A
21	95	7.0	2125	9	US-09-919-172-29	Sequence 29, Appl
22	95	7.0	2649	14	US-10-205-219-169	Sequence 169, App
23	95	7.0	2649	15	US-10-341-434-220	Sequence 220, App
24	95	7.0	2649	15	US-10-341-434-230	Sequence 230, App
c 25	93.5	6.9	408	16	US-10-437-963-160944	Sequence 160944,
26	93.5	6.9	1196	12	US-10-282-122A-52737	Sequence 52737, A
27	93	6.9	610	12	US-10-282-122A-47217	Sequence 47217, A
28	92.5	6.9	430	15	US-10-369-493-5229	Sequence 5229, Ap
29	92.5	6.9	996	9	US-09-815-242-5251	Sequence 5251, Ap
30	92.5	6.9	1009	9	US-09-815-242-12141	Sequence 12141, A
31	92	6.8	889	16	US-10-437-963-197045	Sequence 197045,
32	91.5	6.8	241	12	US-10-412-699B-682	Sequence 682, App
33	91.5	6.8	1847	15	US-10-369-493-1075	Sequence 1075, Ap
34	90.5	6.7	659	12	US-10-424-599-222059	Sequence 222059,
35	90.5	6.7	1163	15	US-10-452-024-107	Sequence 107, App
36	90	6.7	573	12	US-10-425-114-43207	Sequence 43207, A
37	90	6.7	744	16	US-10-437-963-122954	Sequence 122954,
38	90	6.7	820	12	US-10-221-278-590	Sequence 590, App
39	90	6.7	820	15	US-10-291-172-590	Sequence 590, App
40	90	6.7	1089	14	US-10-032-585-7664	Sequence 7664, Ap
41	89.5	6.6	281	15	US-10-289-762-749	Sequence 749, App
c 42	89.5	6.6	460	12	US-10-425-114-63971	Sequence 63971, A
43	89.5	6.6	560	14	US-10-253-904-50	Sequence 50, Appl
44	89.5	6.6	869	12	US-10-282-122A-58561	Sequence 58561, A
45	89	6.6	384	12	US-10-260-708-60	Sequence 60, Appl

ALIGNMENTS

RESULT 1
US-09-732-091-4
; Sequence 4, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; TITLE OF INVENTION: thereof
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-4

Alignment Scores:
Pred. No.: 7.22e-123 Length: 253
Score: 1279.00 Matches: 253

```
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            94.81%         Indels:        0
DB:                     9              Gaps:          0

US-09-732-091-3 (1-759) x US-09-732-091-4 (1-253)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              |||||||||||||||||||||||||||||||||||||
Db        241 AlaAsnGlyAspLysLysSerLeuGlnIleGluSerIle 253

RESULT 2
US-09-732-091-44
; Sequence 44 Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
;  APPLICANT:  Tian, Jing-Hui
;  APPLICANT:  Walker, Richard I.
;  APPLICANT:  Jackson, W. James
;  TITLE OF INVENTION:  Helicobacter proteins, gene sequences and uses
;  TITLE OF INVENTION:  thereof
;  FILE REFERENCE:  7969-088
;  CURRENT APPLICATION NUMBER: US/09/732,091
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 44
;  LENGTH: 265
;  TYPE: PRT
;  ORGANISM: Helicobacter sp.
US-09-732-091-44

Alignment Scores:
Pred. No.:              7.31e-123      Length:        265
Score:                  1279.00        Matches:       253
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            94.81%         Indels:        0
DB:                     9              Gaps:          0

US-09-732-091-3 (1-759) x US-09-732-091-44 (1-265)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 32

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 52

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         53 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 72

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 92

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 112

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        113 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 132

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 152

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 172

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 192

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 212

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 232

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 252

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              |||||||||||||||||||||||||||||||||||||
Db        253 AlaAsnGlyAspLysLysSerLeuGlnIleGluSerIle 265

RESULT 3
US-10-335-977-9162
; Sequence 9162 Application US/10335977
; Publication No. US20040052799A1
```

```
;  GENERAL INFORMATION:
;       APPLICANT: DOUGLAS SMITH et al
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
;                           RELATING TO HELICOBACTER PYLORI FOR
;                           DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 10031
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: LAHIVE & COCKFIELD
;            STREET: 28 State Street
;            CITY: Boston
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02109-1875
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: IBM PC Compatible
;            OPERATING SYSTEM: Windows NT 4.0
;            SOFTWARE: UNIX
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/335,977
;            FILING DATE: 30-Dec-2002
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/993,002
;            FILING DATE: 17-DEC-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mandragouras, Amy E.
;            REGISTRATION NUMBER: 36,207
;            REFERENCE/DOCKET NUMBER: GTN-018
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617)227-7400
;            TELEFAX: (617)742-4214
;  INFORMATION FOR SEQ ID NO: 9162:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 253 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
;       ORIGINAL SOURCE:
;            ORGANISM: Helicobacter pylori
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...253
;       SEQUENCE DESCRIPTION: SEQ ID NO: 9162:
US-10-335-977-9162

Alignment Scores:
Pred. No.:                6.11e-122       Length:          253
Score:                    1270.00         Matches:         251
Percent Similarity:       99.60%          Conservative:    1
Best Local Similarity:    99.21%          Mismatches:      1
Query Match:              94.14%          Indels:          0
DB:                       12              Gaps:            0

US-09-732-091-3 (1-759) x US-10-335-977-9162 (1-253)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              ||||||  ||||||||||||||||||||||||:::
Db        241 AlaAsnGluAspLysLysSerLeuGlnIleGluSerVal 253

RESULT 4
US-10-335-977-9163
; Sequence 9163, Application US/10335977
; Publication No. US20040052799A1
;  GENERAL INFORMATION:
;       APPLICANT: DOUGLAS SMITH et al
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
;                           RELATING TO HELICOBACTER PYLORI FOR
;                           DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 10031
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: LAHIVE & COCKFIELD
;            STREET: 28 State Street
;            CITY: Boston
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02109-1875
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: IBM PC Compatible
;            OPERATING SYSTEM: Windows NT 4.0
;            SOFTWARE: UNIX
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/335,977
;            FILING DATE: 30-Dec-2002
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/993,002
;            FILING DATE: 17-DEC-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mandragouras, Amy E.
;            REGISTRATION NUMBER: 36,207
;            REFERENCE/DOCKET NUMBER: GTN-018
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617)227-7400
;            TELEFAX: (617)742-4214
;  INFORMATION FOR SEQ ID NO: 9163:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 253 amino acids
```

```
;           TYPE: amino acid
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
;       ORIGINAL SOURCE:
;           ORGANISM: Helicobacter pylori
;       FEATURE:
;           NAME/KEY:  misc_feature
;           LOCATION: (B) LOCATION 1...253
;       SEQUENCE DESCRIPTION: SEQ ID NO: 9163:
US-10-335-977-9163

Alignment Scores:
Pred. No.:                6.11e-122      Length:          253
Score:                    1270.00        Matches:         251
Percent Similarity:       99.60%         Conservative:    1
Best Local Similarity:    99.21%         Mismatches:      1
Query Match:              94.14%         Indels:          0
DB:                       12             Gaps:            0

US-09-732-091-3 (1-759) x US-10-335-977-9163 (1-253)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 40

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 60

Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 80

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 100

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 120

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 140

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 160

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 180

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 200

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 220

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 240

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
              |||||||  ||||||||||||||||||||||||||:::
Db        241 AlaAsnGluAspLysLysSerLeuGlnIleGluSerVal 253
```

```
RESULT 5
US-10-335-977-9164
; Sequence 9164, Application US/10335977
; Publication No. US20040052799A1
;   GENERAL INFORMATION:
;       APPLICANT: DOUGLAS SMITH et al
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
;                           RELATING TO HELICOBACTER PYLORI FOR
;                           DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 10031
;       CORRESPONDENCE ADDRESS:
;           ADDRESSEE: LAHIVE & COCKFIELD
;           STREET: 28 State Street
;           CITY: Boston
;           STATE: Massachusetts
;           COUNTRY: USA
;           ZIP: 02109-1875
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE: CD/ROM ISO9660
;           COMPUTER: IBM PC Compatible
;           OPERATING SYSTEM: Windows NT 4.0
;           SOFTWARE: UNIX
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/10/335,977
;           FILING DATE: 30-Dec-2002
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: 08/993,002
;           FILING DATE: 17-DEC-1997
;       ATTORNEY/AGENT INFORMATION:
;           NAME: Mandragouras, Amy E.
;           REGISTRATION NUMBER: 36,207
;           REFERENCE/DOCKET NUMBER: GTN-018
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (617)227-7400
;           TELEFAX: (617)742-4214
;   INFORMATION FOR SEQ ID NO: 9164:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 256 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
;       ORIGINAL SOURCE:
;           ORGANISM: Helicobacter pylori
;       FEATURE:
;           NAME/KEY:  misc_feature
;           LOCATION: (B) LOCATION 1...256
;       SEQUENCE DESCRIPTION: SEQ ID NO: 9164:
US-10-335-977-9164

Alignment Scores:
Pred. No.:                6.13e-122      Length:          256
Score:                    1270.00        Matches:         251
Percent Similarity:       99.60%         Conservative:    1
Best Local Similarity:    99.21%         Mismatches:      1
Query Match:              94.14%         Indels:          0
DB:                       12             Gaps:            0

US-09-732-091-3 (1-759) x US-10-335-977-9164 (1-256)

Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 23

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 LeuAspLeuPheGluValLeuValPheGlyLysAspGlyGluLysArgHisAsnGluLys 43

Qy        121 CTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         44 LeuThrSerSerIleGluTyrLysArgHisGlyAspAspTyrAlaLysTyrAlaGluArg 63
```

```
Qy        181 ATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAA 240
Db         64 IleAlaGluGluLeuGlnTyrTyrGlySerAsnSerPheAlaSerPheIleLysGlyGlu 83

Qy        241 GGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAAC 300
Db         84 GlyValLeuTyrLysGluIleLeuCysAspValCysAspLysLeuLysValAsnTyrAsn 103

Qy        301 AAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGT 360
Db        104 LysLysThrGluThrThrLeuIleGluGlnAsnMetLeuSerLysIleLeuGluArgSer 123

Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
Db        124 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 143

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGG 480
Db        144 ThrAspAsnLeuAsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGly 163

Qy        481 GGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTA 540
Db        164 GlyPheLysSerTyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeu 183

Qy        541 GGGCGTGGTTTATCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
Db        184 GlyArgGlyLeuSerLeuAlaGlyAsnGlnValLeuThrArgThrLeuSerPheLeuThr 203

Qy        601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
Db        204 GlyProValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAla 223

Qy        661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
Db        224 TyrArgValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGln 243

Qy        721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
Db        244 AlaAsnGluAspLysLysSerLeuGlnIleGluSerVal 256
```

```
RESULT 6
US-10-335-977-7699
; Sequence 7699, Application JS/10335977
; Publication No. US20040052799A1
;   GENERAL INFORMATION:
;       APPLICANT: DOUGLAS SMITH et al
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
;                           RELATING TO HELICOBACTER PYLORI FOR
;                           DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 10031
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: LAHIVE & COCKFIELD
;            STREET: 28 State Street
;            CITY: Boston
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02109-1875
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: IBM PC Compatible
;            OPERATING SYSTEM: Windows NT 4.0
;            SOFTWARE: UNIX
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/335,977
;            FILING DATE: 30-Dec-2002
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/993,002
;            FILING DATE: 17-DEC-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mandragouras, Amy E.
;            REGISTRATION NUMBER: 36,207
;            REFERENCE/DOCKET NUMBER: GTN-018
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617)227-7400
;            TELEFAX: (617)742-4214
;   INFORMATION FOR SEQ ID NO: 7699:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 248 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
;       ORIGINAL SOURCE:
;            ORGANISM: Helicobacter pylori
;       FEATURE:
;            NAME/KEY: misc_feature
;            LOCATION: (B) LOCATION 1...248
;       SEQUENCE DESCRIPTION: SEQ ID NO: 7699:
US-10-335-977-7699

Alignment Scores:
Pred. No.:                1.9e-65        Length:          248
Score:                    722.00         Matches:         141
Percent Similarity:       75.64%         Conservative:    36
Best Local Similarity:    60.26%         Mismatches:      53
Query Match:              53.52%         Indels:          4
DB:                       12             Gaps:            2

US-09-732-091-3 (1-759) x US-10-335-977-7699 (1-248)

Qy         22 GACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTT 81
Db         11 AspLeuGluPheLeuLysArgLeuSerSerSerAspLeuLysAspLeuPheAspAlaLeu 30

Qy         82 GTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATAC 141
Db         31 ValTyrAspGluAspGlyThrLeuArgMetAsnGluGluLeuThrSerLeuThrGluTyr 50

Qy        142 AAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATAC 201
Db         51 GlnArgTyrGlyHisAspTyrAlaLysTyrProArgArgIleAlaGluGluLeuGlnArg 70

Qy        202 TATGGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATT 261
Db         71 TyrGlyGlyAsnSerPheAlaAsnPhePheArgAspGluGlyValLeuTyrLysGluIle 90

Qy        262 TTATGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTA 321
Db         91 LeuCysAspAlaCysAspHisLeuAspIleAsnTyrAsnGluArgSerAlaThrSerLeu 110

Qy        322 ATTGAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAA 381
Db        111 IleGluGlnAsnMetLeuSerLysLeuLeuLysAspSerLeuGluLysMetSerGlyArg 130

Qy        382 GAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAACACGGACAATTTA--------- 432
Db        131 GluIleLysGluLeuCysAspGlyLeuGlyMetProAsnIleAspLysValIleGlyGlu 150

Qy        433 AACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCT 492
Db        151 AsnLysGlnValLeuIleAlaSerValLeuThrLeuPheLysAlaGlyGlySerHisSer 170

Qy        493 TATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTA 552
Db        171 TyrAlaLeuAlaValAlaValAlaAspAlaMetValArgGlnThrLeuGlyHisGlyLeu 190

Qy        553 ---TCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTT 609
Db        191 SerSerValValGlyLysValAlaLeuLysLysThrLeuAspIleLeuAlaGlyProIle 210

Qy        610 GGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTA 669
Db        211 GlyTrpValIleThrGlyAlaLeuValSerIleAsnLeuAlaGlyProAlaTyrArgVal 230

Qy        670 ACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAA 711
```

```
Db        231 ThrValProAlaCysValLeuValAlaThrLeuArgLysLys 244

RESULT 7
US-10-335-977-7698
; Sequence 7698, Application US/10335977
; Publication No. US20040052799A1
;   GENERAL INFORMATION:
;        APPLICANT: DOUGLAS SMITH et al
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
;                            RELATING TO HELICOBACTER PYLORI FOR
;                            DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 10031
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: LAHIVE & COCKFIELD
;             STREET: 28 State Street
;             CITY: Boston
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02109-1875
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: IBM PC Compatible
;             OPERATING SYSTEM: Windows NT 4.0
;             SOFTWARE: UNIX
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/335,977
;             FILING DATE: 30-Dec-2002
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/993,002
;             FILING DATE: 17-DEC-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mandragouras, Amy E.
;             REGISTRATION NUMBER: 36,207
;             REFERENCE/DOCKET NUMBER: GTN-018
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617)227-7400
;             TELEFAX: (617)742-4214
;   INFORMATION FOR SEQ ID NO: 7698:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 237 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Helicobacter pylori
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...237
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7698:
US-10-335-977-7698

Alignment Scores:
Pred. No.:              7.79e-65      Length:          237
Score:                  716.00        Matches:         140
Percent Similarity:     75.54%        Conservative:     36
Best Local Similarity:  60.09%        Mismatches:       53
Query Match:            53.08%        Indels:            4
DB:                     12            Gaps:              2

US-09-732-091-3 (1-759) x US-10-335-977-7698 (1-237)

Qy         25 TTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTT 84
              ||||||||||||||:::|||   ||||||||||||   |||||||||:::   ||||||
Db          1 LeuGluPheLeuLysArgLeuSerSerSerAspLeuLysAspLeuPheAspAlaLeuVal 20

Qy         85 TTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTGACCAGCTCCATAGAATACAAA 144
              :::   :::||||||      |||   ||||||:::||||||||      ||||||:::
Db         21 TyrAspGluAspGlyThrLeuArgMetAsnGluGluLeuThrSerLeuThrGluTyrGln 40

Qy        145 AGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTAT 204
              |||:::|||   |||||||||||||||      ||||||||||||||||||||   |||
Db         41 ArgTyrGlyHisAspTyrAlaLysTyrProArgArgIleAlaGluGluLeuGlnArgTyr 60

Qy        205 GGGAGCAATAGTTTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTA 264
              |||   ||||||||||||:::|||   :::   ||||||||||||||||||||||||||
Db         61 GlyGlyAsnSerPheAlaAsnPhePheArgAspGluGlyValLeuTyrLysGluIleLeu 80

Qy        265 TGCGATGTGTGCGATAAATTAAAGGTCAATTACAACAAGAAAACTGAAACGACTTTAATT 324
              ||||||   ||||||   |||   :::||||||||:::::::::   |||:::||||||
Db         81 CysAspAlaCysAspHisLeuAspIleAsnTyrAsnGluArgSerAlaThrSerLeuIle 100

Qy        325 GAACAAAACATGCTTTCTAAAATCTTAGAAAGAAGTTTGGAAGAAATGGATGATGAAGAA 384
              |||||||||||||||||||:::|||:::   |||||||||:::|||          |||
Db        101 GluGlnAsnMetLeuSerLysLeuLeuLysAspSerLeuGluLysMetSerGlyArgGlu 120

Qy        385 GTGAAAGAAATGTGCGATGAATTATCCATAAAAAACACGGACAATTTA---------AAC 435
              :::|||||:::||||||   |||   :::   |||   |||   :::          |||
Db        121 IleLysGluLeuCysAspGlyLeuGlyMetProAsnIleAspLysValIleGlyGluAsn 140

Qy        436 AGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTAT 495
              :::|||   |||   |||:::   ||||||||||||||   ||||||   ||||||
Db        141 LysGlnValLeuIleAlaSerValLeuThrLeuPheLysAlaGlyGlySerHisSerTyr 160

Qy        496 CAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTA--- 552
              |||||||||   ||||||:::|||:::   :::     ||||||   ||||||
Db        161 AlaLeuAlaValAlaValAlaAspAlaMetValArgGlnThrLeuGlyHisGlyLeuSer 180

Qy        553 TCGCTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACAGGTCCTGTTGGC 612
              |||:::   |||           |||   :::||||||      |||   ||||||:::|||
Db        181 SerValValGlyLysValAlaLeuLysLysThrLeuAspIleLeuAlaGlyProIleGly 200

Qy        613 TGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGGGTAACC 672
              |||:::||||||||||          :::|||:::::|||||||||||||||||||||||
Db        201 TrpValIleThrGlyAlaLeuValSerIleAsnLeuAlaGlyProAlaTyrArgValThr 220

Qy        673 ATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAA 711
              :::|||||||||:::::|||||||||||||||   |||
Db        221 ValProAlaCysValLeuValAlaThrLeuArgLysLys 233

RESULT 8
US-09-882-227-414
; Sequence 414, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
;  APPLICANT: Kleanthous, Harold
;  APPLICANT:  Al-Garawi, Amal
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Tomb, Jean-Francois
;  APPLICANT:  Ooomen, Raymond P.
;  TITLE OF INVENTION: Identification of Polynucleotides
;  TITLE OF INVENTION:  Encoding No. US20030158396A1el Helicobacter Polypeptides in t
;  TITLE OF INVENTION:  Genome
;  FILE REFERENCE: 06132/047002
;  CURRENT APPLICATION NUMBER: US/09/882,227
;  CURRENT FILING DATE:  2001-06-15
;  PRIOR APPLICATION NUMBER: US 08/902,615
;  PRIOR FILING DATE: 1997-07-29
;  NUMBER OF SEQ ID NOS: 638
;  SOFTWARE: PastSEQ for Windows Version 4.0
; SEQ ID NO 414
;   LENGTH: 155
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-09-882-227-414

Alignment Scores:
Pred. No.:              8.85e-40      Length:          155
Score:                  472.50        Matches:          93
Percent Similarity:     78.23%        Conservative:     22
Best Local Similarity:  63.27%        Mismatches:       29
Query Match:            35.03%        Indels:            3
DB:                     10            Gaps:              1

US-09-732-091-3 (1-759) x US-09-882-227-414 (1-155)
```

```
Qy        112 AATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATAC 171
              ||||||  |||||:::|||  ||||||||||||:::||  ||||||||||||||||
Db          2 AsnGluAspLeuThrAsnSerThrGluTyrLysArgTyrGlyHisAspTyrAlaLysTyr 21

Qy        172 GCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATT 231
              ||||||||||||||||||:::||||||  ||||||||||||:::|||
Db         22 ProArgArgIleAlaGluGluLeuGlnHisTyrGlyGlyAsnSerPheAlaAsnPhePhe 41

Qy        232 AAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTC 291
              :::  ||||||||||||||||||||||||||||||  ||||||  |||||||||
Db         42 ArgAspGluGlyValLeuTyrLysGluIleLeuCysAspAlaCysAspHisLeuLysVal 61

Qy        292 AATTACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTA 351
              ||||||||:::::::::  |||:::||||||||||||||||||||||:::|||
Db         62 AsnTyrAsnGluGluSerAlaThrSerLeuIleGluGlnAsnMetLeuSerLysLeuLeu 81

Qy        352 GAAAGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCC 411
              :::  |||||||||:::|||  |||:::||||||:::|||:::||||||
Db         82 LysAspSerLeuGluLysMetSerArgArgGluIleLysGluLeuCysAsnGluLeuGly 101

Qy        412 ATAAAAAACACGGACAATTTA---------AACAGACAAGCCTTAAGCGCGGCGACTTTA 462
              :::  |||  |||  :::       |||:::|||  |||  |||:::||||||
Db        102 MetThrAsnIleAspLysValIleGlyGluAsnLysGlnValLeuIleAlaSerThrLeu 121

Qy        463 ACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCTGTCATTGTTGCGAATGCG 522
              ||||||||||||  ||||||  ||||||  ||||||||  ||||||:::|||
Db        122 ThrLeuPheLysAlaGlyGlySerHisSerTyrAlaLeuAlaValSerValAlaAspAla 141

Qy        523 GTCGCAAAAACCATTCTAGGG 543
              :::  :::  ||||||
Db        142 MetValArgGlnThrLeuGly 148
```

RESULT 9
US-09-732-091-20
; Sequence 20, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; TITLE OF INVENTION: thereof
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-20

Alignment Scores:

Pred. No.:	3.55e-17	Length:	49
Score:	252.00	Matches:	49
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	18.68%	Indels:	0
DB:	9	Gaps:	0

US-09-732-091-3 (1-759) x US-09-732-091-20 (1-49)

```
Qy        607 GTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCTTATAGG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ValGlyTrpIleIleThrGlyValTrpThrAlaIleAspIleAlaGlyProAlaTyrArg 20

Qy        667 GTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAAGCCAAT 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ValThrIleProAlaCysIleValValAlaThrLeuArgLeuLysThrGlnGlnAlaAsn 40

Qy        727 GGAGATAAGAAGTCGTTGCAAATAGAA 753
              |||||||||||||||||||||||||||
```

```
Db         41 GlyAspLysLysSerLeuGlnIleGlu 49
```

RESULT 10
US-09-732-091-17
; Sequence 17, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; TITLE OF INVENTION: thereof
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-17

Alignment Scores:

Pred. No.:	1.23e-11	Length:	38
Score:	198.00	Matches:	38
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	14.68%	Indels:	0
DB:	9	Gaps:	0

US-09-732-091-3 (1-759) x US-09-732-091-17 (1-38)

```
Qy        157 GATTACGCTAAATACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGT 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AspTyrAlaLysTyrAlaGluArgIleAlaGluGluLeuGlnTyrTyrGlySerAsnSer 20

Qy        217 TTTGCGAGTTTCATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGAT 270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaSerPheIleLysGlyGluGlyValLeuTyrLysGluIleLeuCysAsp 38
```

RESULT 11
US-09-732-091-19
; Sequence 19, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; TITLE OF INVENTION: thereof
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-19

Alignment Scores:

Pred. No.:	3.23e-11	Length:	41
Score:	194.00	Matches:	41
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	14.38%	Indels:	0
DB:	9	Gaps:	0

US-09-732-091-3 (1-759) x US-09-732-091-19 (1-41)

```
Qy        433 AACAGACAAGCCTTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCT 492
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AsnArgGlnAlaLeuSerAlaAlaThrLeuThrLeuPheLysMetGlyGlyPheLysSer 20

Qy        493 TATCAATTAGCTGTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTA 552
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 TyrGlnLeuAlaValIleValAlaAsnAlaValAlaLysThrIleLeuGlyArgGlyLeu 40

Qy        553 TCG 555
                |||
Db         41 Ser 41
```

RESULT 12
US-09-732-091-18
; Sequence 18, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; TITLE OF INVENTION: thereof
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-18

Alignment Scores:

Pred. No.:	6.37e-07	Length:	30
Score:	152.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	11.27%	Indels:	0
DB:	9	Gaps:	0

US-09-732-091-3 (1-759) x US-09-732-091-18 (1-30)

```
Qy        361 TTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATAAAAAAC 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGluGluMetAspAspGluGluValLysGluMetCysAspGluLeuSerIleLysAsn 20

Qy        421 ACGGACAATTTAAACAGACAAGCCTTAAGC 450
                ||||||||||||||||||||||||||||||
Db         21 ThrAspAsnLeuAsnArgGlnAlaLeuSer 30
```

RESULT 13
US-09-732-091-16
; Sequence 16, Application US/09732091
; Patent No. US20020107368A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Jing-Hui
; APPLICANT: Walker, Richard I.
; APPLICANT: Jackson, W. James
; TITLE OF INVENTION: Helicobacter proteins, gene sequences and uses
; TITLE OF INVENTION: thereof
; FILE REFERENCE: 7969-088
; CURRENT APPLICATION NUMBER: US/09/732,091
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Helicobacter sp.
US-09-732-091-16

Alignment Scores:

Pred. No.:	1.02e-06	Length:	30
Score:	150.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	11.12%	Indels:	0
DB:	9	Gaps:	0

US-09-732-091-3 (1-759) x US-09-732-091-16 (1-30)

```
Qy          1 ATGGCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTA 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaTyrLysTyrAspArgAspLeuGluPheLeuLysGlnLeuGluSerSerAspLeu 20

Qy         61 TTGGATTTGTTTGAGGTGCTTGTTTTTGGT 90
                ||||||||||||||||||||||||||||||
Db         21 LeuAspLeuPheGluValLeuValPheGly 30
```

RESULT 14
US-10-282-122A-43832
; Sequence 43832, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43832
; LENGTH: 1009
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-43832

Alignment Scores:

Pred. No.:	0.515	Length:	1009
Score:	98.50	Matches:	66
Percent Similarity:	37.88%	Conservative:	45
Best Local Similarity:	22.53%	Mismatches:	83
Query Match:	7.30%	Indels:	99
DB:	12	Gaps:	16

```
US-09-732-091-3 (1-759) x US-10-282-122A-43832 (1-1009)

Qy          13 TATGATAGAGACTTGGAATTT----------------------TTAAAGCAATTGGAA 48
               |||:::   |||||||||                      :::   :::|||
Db         708 TyrAsnAsnAspLeuGluLeuAsnGluGlnSerIleGluMetGluMetSerArgLeuAsn 727

Qy          49 TCTAGTGATTTATTGGATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGA 108
                  :::|||      |||:::  |||::::::  :::      :::  |||:::
Db         728 LeuThrAspAspAsnAspIleAsnGluIleIleAlaTrpArgGlyGluGlnGluGlu--- 746

Qy         109 CACAATGAAAAACTGACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAA 168
                                  :::|||  ||||||     |||  |||  |||
Db         747 ------------------------LeuGluGlnLysArg------AspThrTyrLysLys 756

Qy         169 TACGCAGAAAGAATCGCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTC 228
                                 |||:::                              |||
Db         757 ArgTyrHisGluPheGluMetGluIleAla---------------ArgLeuGluSerLeu 771

Qy         229 ATTAAAGGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAG 288
                  |||                 |||||:::|||           ||||||||||||
Db         772 ThrLysAsp--------------LysGluLeuLeuAspSer------AspLysLeuLys 784

Qy         289 GTCAATTACAACAAGAAAACTGAA------ACGACTTTAATTGAA--------------- 327
                     |||   :::|||   |||        |||||||||:::
Db         785 AspGluTyrGluGlnLysLysGluLysMetAsnThrLeuIleAspGluTyrSerAlaVal 804

Qy         328 --------------------------------CAAAACATGCTTTCTAAAATCTTAGAA 354
                                               |||:::::::::|||   |||
Db         805 HisTyrGlnCysGlnAsnAsnIleAsnLysThrGlnSerIleValSerHisIleAsnTyr 824

Qy         355 AGAAGTTTGGAAGAAATGGATGATGAAGAAGTGAAAGAAATGTGCGATGAATTATCCATA 414
                  :::   |||      |||   :::|||:::   ::::::   :::   :::|||
Db         825 LeuAsnGlnGluLeuLysAspGlnGlnGluIlePheGlnLeuAlaGluIleValSerGly 844

Qy         415 AAAAACACGGACAATTTAAAC----------------------------AGACAAGCC 444
               ||||||     ||||||                               |||
Db         845 LysAsnAsnLysAsnLeuThrLeuGluAsnPheValLeuIleTyrTyrLeuAspGlnIle 864

Qy         445 TTAAGCGCGGCGACTTTAACGCTGTTTAAAATGGGGGGTTTTAAATCTTATCAATTAGCT 504
               ::::::   |||   |||   |||      |||                |||||||||
Db         865 IleAlaGlnAlaAsnLeuArgLeuAlaThrMetSerAsp---AsnArgTyrGlnLeu--- 882

Qy         505 GTCATTGTTGCGAATGCGGTCGCAAAAACCATTCTAGGGCGTGGTTTATCG--------- 555
                              :::   :::      :::      |||||||||
Db         883 ------------------IleArgArgGluAlaValSerHisGlyLeuSerGlyLeuGlu 896

Qy         556 ---------------CTTGCGGGCAATCAGGTGCTTACAAGAACTCTGAGCTTTTTAACA 600
                           |||      |||:::      :::|||   :::|||   |||:::
Db         897 IleAspValPheAspLeuHisSerAsnLys------SerArgHisIleSerSerLeuSer 914

Qy         601 GGTCCTGTTGGCTGGATCATTACAGGCGTATGGACAGCGATTGATATTGCAGGGCCGGCT 660
               |||                                                   |||
Db         915 Gly-------------------------------------------------GlyGluThr 918

Qy         661 TATAGGGTAACCATACCGGCATGCATTGTGGTTGCCACTTTACGCCTAAAAACACAGCAA 720
               ::::::   ::::::         :::   ::::::   :::           ||||||
Db         919 PheGlnSerSerLeuAlaLeuAlaLeuGlyLeuSerGluIle---------ValGlnGln 935

Qy         721 GCCAATGGAGATAAGAAGTCGTTGCAAATAGAATCCATT 759
                  :::|||         :::   :::|||||||||
Db         936 GlnSerGly---------GlyIleSerLeuGluSerIle 945

RESULT 15
US-10-437-963-191043
; Sequence 191043, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated Wit
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191043
;   LENGTH: 815
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_8739C.1.pep
US-10-437-963-191043

Alignment Scores:
Pred. No.:                0.618          Length:          815
Score:                    97.50          Matches:         45
Percent Similarity:       44.38%         Conservative:    26
Best Local Similarity:    28.12%         Mismatches:      62
Query Match:              7.23%          Indels:          27
DB:                       16             Gaps:            8

US-09-732-091-3 (1-759) x US-10-437-963-191043 (1-815)

Qy           4 GCATACAAATATGATAGAGACTTGGAATTTTTAAAGCAATTGGAATCTAGTGATTTATTG 63
               |||   :::   ::::::|||||||                |||:::   |||   |||
Db         531 AlaAsnArgMetAsnGlnAspLeu------------ValLeuGlnIleSerLysLeuGln 546

Qy          64 GATTTGTTTGAGGTGCTTGTTTTTGGTAAAGACGGCGAAAAAAGACACAATGAAAAACTG 123
               |||   |||      :::      |||          :::              |||   :::
Db         547 AspGluPheAsnGluMetGlnGluGlyLeuThrAsnLysLeuGlyGluValGluSerVal 566

Qy         124 ACCAGCTCCATAGAATACAAAAGGCATGGCGATGATTACGCTAAATACGCAGAAAGAATC 183
                  :::   ::::::                    |||:::              |||   :::
Db         567 SerLysAlaLeu---------------SerAspGluLeuValSerValLysGluMetVal 581

Qy         184 ---------GCTGAAGAGTTGCAATACTATGGGAGCAATAGTTTTGCGAGTTTCATTAAA 234
                        |||||||||:::          ||||||      ||||||   ::::::
Db         582 HisLysGlyGlnGluGluLeuGlu---AlaThrSerAsnGluLeuAlaSerIleValGlu 600

Qy         235 GGCGAAGGAGTCTTATACAAAGAGATTTTATGCGATGTGTGCGATAAATTAAAGGTCAAT 294
                     |||   ||||||:::|||   ||||||      ||||||
Db         601 AlaArgAspAsnLeuLysLysGluLeuLeu---AspValPheLysLysLeu--------- 616

Qy         295 TACAACAAGAAAACTGAAACGACTTTAATTGAACAAAACATGCTTTCTAAAATCTTAGAA 354
                  :::   |||          |||:::::::::                      |||
Db         617 ------GluSerThrSerGlnGluLeuValAspGluArgLysThrValThrThrLeuAsn 634

Qy         355 AGAAGTTTGGAA------------------GAAATGGATGATGAAGAAGTGAAAGAAATG 396
               |||   ||||||                :::||||||   |||      |||   :::
Db         635 ArgGluLeuGluAlaLeuValLysGlnLeuGlnMetAspSerGluAlaArgLysAlaLeu 654

Qy         397 TGCGATGAATTA------TCCATAAAAAACACGGACAATTTAAACAGACAAGCCTTAAGC 450
                  :::|||      :::  |||:::   |||   :::||||||   |||||||||
Db         655 GluAlaAspLeuAspGluAlaThrLysSerLeuAspGluMetAsnArgSerAlaLeuSer 674


Search completed: July  5, 2004, 03:10:40
Job time : 71.5 secs
```